Analyzing Timber Trade in Brazil: assessing timber networks and supply chains

Victor Russo,[†] Bernardo Costa,[†] Felipe Moreno-Vera,[‡] Guilherme Toledo,[†] Osni Brito de Jesus,[†] Robson Vieira,[¶] Marco Lentini,[§] Jorge Poco,[‡] and Luis Gustavo Nonato[*,†]

†*University of São Paulo, ICMC-USP, Brazil*

‡*Fundação Getúlio Vargas, EMAp-FGV, Brazil*

¶*RFV Consultancy and Systems, Brazil*

§*Imaflora, Brazil*

E-mail: gnonato@icmc.usp.br

Abstract

Brazilian Amazon forest degradation is driven by factors such as fire, mining, and notably, illegal logging. The Brazilian government has implemented control mechanisms to combat illegal timber extraction that have positively impacted deforestation rates. Under these regulations, any wood product, from raw logs to processed lumber, must be registered in control systems before transportation. This allows for analyzing the volume of wood products moving between each pair of timber companies within a specified timeframe. A challenge in this context is the existence of three different and only partially integrated control systems, making the comprehensive analysis of the entire timber market difficult. In this work, we integrate timber transportation data from the three different systems, enabling the creation of what we refer to as Timber Trade Networks (TTNs). From the TTNs, one can identify companies or groups of companies that operate contrary to expected standards, raising suspicions about their compliance with legal regulations. Furthermore, we propose a method to compute probable supply chains of timber companies, addressing a critical traceability challenge that has long concerned the Brazilian government. Among the results, we show that certain TTNs 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 significantly enhance customer confidence in the legality of purchased timber products.

Introduction

Brazilian Amazon forest degradation has long been a source of great concern for environmental agencies in Brazil and abroad due to its impact on biodiversity,[1] environment,[2,3] and even on the control of infectious diseases.[4] Forest degradation can be caused by factors such as fire[5] and mining,[6] but mainly by illegal logging.[7] Aiming to reduce illegal timber extraction and, thus, forest degradation, in 2006, the Brazilian government implemented a computational system for controlling the legal logging activity,[8] with a positive impact on deforestation, although illegal logging continues to occur. The control system is divided into three not fully integrated subsystems: the SINAFLOR, a nationwide system that covers most Brazilian states except the states of Mato Grosso and Pará, which have their own control systems, the SISFLORA-MT and SISFLORA-PA, respectively. For SINAFLOR, the control is carried out through the Document of Forest Origin (Documento de Origem Florestal - DOF), while SISFLORA-MT and SISFLORA-PA rely on a similar document called Forest Guide (Guia Florestal - GF). Any wood product, from raw logs to processed lumber, must be registered in a GF or a DOF to be transported, even between branches of the same company.

The information in each DOF and GF includes the volume and species being transported, origin and destination georeferencing, transportation date, and data about the origin and destination companies. Therefore, from the GFs and DOFs it is possible to generate what we term *Timber Trade Network* (TTN), with nodes corresponding to companies operating in the timber market and edges linking pairs of companies that traded timber. In particular, a timber trade network can be built for each wood species. For a given species, the volume of timber flowing through each edge of the corresponding TTN can also be computed from the GFs and DOFs, allowing to estimate the companies' "mass balance" as the ratio between outflow and inflow volumes.

Timber is expected to enter a Timber Trade Network through nodes representing *licensed forests*, which are authorized logging areas explicitly 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 edges of a TTN 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 Timber Trade Networks 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 probable *supply chains*. By identifying these potential timber routes, it becomes possible to assess the operational scale of businesses within timber supply chains, determine the volume transported through these chains, and pinpoint the likely forest origins of the wood. Therefore, analyzing probable supply chains enables more confident decision-making regarding the legality of the traded timber. Additionally, by comparing the volume of timber entering a supply chain from licensed forests with the registered volumes (declared in the GFs and DOFs) flowing through the chain we can identify disparities that may indicate illegitimate activities.

We have compiled and integrated data from SINAFLOR, SISFLORA-MT, and SISFLORA-PA from 2010 to 2020 to execute the aforementioned analysis. This data fusion enabled the construction of Timber Trade Networks for a selection of high-value hardwood species. Leveraging the timber volumes coursing through these TTNs, we devised a method to calculate the k most probable supply chains of timber companies. This process makes it possible to identify the businesses involved in each chain, including licensed forests the timber is likely to have originated from. We can scrutinize whether the timber volumes traversing the probable chains align with the logging volumes leaving licensed forests. This analysis enables the identification of groups of companies engaged in the timber trade market without being linked to licensed forests. Moreover, by analyzing the mass balance of companies involved in particular supply chains we can find discrepancies that contradict expected norms, indicating potential illegalities.

Our proposed 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 probable timber supply chains. This approach offers a viable alternative to implementing costly and potentially ineffective traceability resources, an avenue currently under consideration by Brazilian inspection agencies.

Material and 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 will collectively refer to both authorization types as *licensed forests*.

We obtained timber transportation data from the national SINAFLOR system, which is reachable from IBAMA's public data portal.[8] Additionally, Imaflora, a non-governmental organization collaborating on this project, acquired data from the SISFLORA-MT and SISFLORA-PA through collaborative agreements with the environmental regulatory bodies of the respective states. The collected data resulted in 31,913,611 DOFs and 18,601,146 GFs. This endeavor has resulted in the compilation of a comprehensive and diverse dataset, rich with information crucial for our analysis.

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 (lat, long). We additionally employed location coordinates to establish a unique identifier for licensed forests, as different areas might be linked to the same exploration authorization. To further complete the dataset, we extracted information from PDF files publicly available in the SISFLORA-PA system, mainly to fill location coordinate missing in several GFs. Beyond licensed forests and timber companies, other important players in the timber trade market are the *end consumers*, in which the transportation flow ultimately terminates. Therefore, entities that appear in the GFs and DOFs only as destination are considered end consumers. In particular, the destinations marked as "Exportation" give rise to an end consumer in the dataset. End consumers account for approximately 85% of the entities involved in the timber trade network. At the end of the integration process describe above, licensed forests, timber companies, and end consumers from the three control systems are uniquely identified as "actors" operating in the timber trade ecosystem.

The transport data has 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. An effort has been made to standardize and aggregate timber products and species names, encompassing scientific and popular nomenclature. The standardization process ensured uniformity in representing wood and product types, facilitating interpretation. 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 status, for instance "canceled" and "active") are also recorded in the dataset. The remaining attributes account for the *autorization number*, *autorization type*, and *registration type*.

In summary, the data compilation process gave rise to a dataset consists 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; 4) a list of forest products.

Timber Trade Networks and Supply Chains

The DOFs and GFs explicitly state the volume of each wood species transported from a licensed forest or timber company to another company or end consumer. Therefore, the timber trade network of a given wood species can be built as a directed graph with three types of nodes: licensed forests, timber trade companies, and end consumers. Nodes corresponding to licensed forests are source nodes, as hardwood (or wood raw products) can only outflow those nodes. Nodes representing timber trade companies have inflow and outflow edges associated with them. End consumers are inflow-only sink nodes.

In mathematical terms, the TTN of a given wood species $\mathcal{S}$ is a directed graph $G_{\mathcal{S}} = \{V, E, W\}$, 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 $\mathcal{S}$ from v_i to v_j. The entries w_{ij} in the weight matrix W encapsulate the total volume of species $\mathcal{S}$ was transported from v_i to v_j between 2010 and 2020.

The mass balance of each node $v_j \in V_c$ can be defined as the ratio between outflow and inflow volumes, i.e., $mb_{v_j} = \sum_i w_{ji} / \sum_i w_{ij}$. In our context, a supply chain associated with a node $v \in V_c \cup V_e$ is a sorted sequence of nodes $sc_v = \{v_{c_0}, v_{c_1}, \cdots, v_{c_s}\}$, where $v_{c_s} = v$, $v_{c_0} \in V_f$ (v_{c_0} is a licensed forest), $v_{c_i} \in V_c$, $i = 1, \ldots, s-1$, $v_{c_i} \neq v_{c_j}$ for all $i \neq j$, with a directed edge from $v_{c_{i-1}}$ to v_{c_i} for $i = 0, \ldots, s$. In other words, a supply chain of a node v is a directed path without loops, starting in a licensed forest and ending in the node v. The total timber volume transported throughout a supply chain is given by $\sigma(sc_v) = \sum_i w_{c_{i-1}c_i}$, $i = 1, \ldots, s$.

Supposedly, timber may only enter a TTN from licensed forest nodes. Timber leaves a TTN in two ways: due to losses inherent to the processing carried out by sawmills or sales to end consumers (including exports). Therefore, the mass balance mb_{v_j} of each node $v_j \in V_c$ should always be smaller or equal to one. Ratios greater than one 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 of time (except in cases where the company possesses an extensive stockpile of wood, which is uncommon).

Most Probable Supply Chains

Given the impossibility of precisely tracing wood products' pathways, 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:

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

If we flip the TTN edges' directions and the multiply their associated volumes by -1 (volumes associated with edges become negative), the problem defined in Eq. (1) can be solved by a variant of Dijkstra's algorithm called *single source shortest-path problem.*[9] Such an algorithm has a sub-polynomial 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 $G_{\mathcal{S}}$ (in the same connected component as v), in particular to licensed forests, thus computing timber supply chains for v.

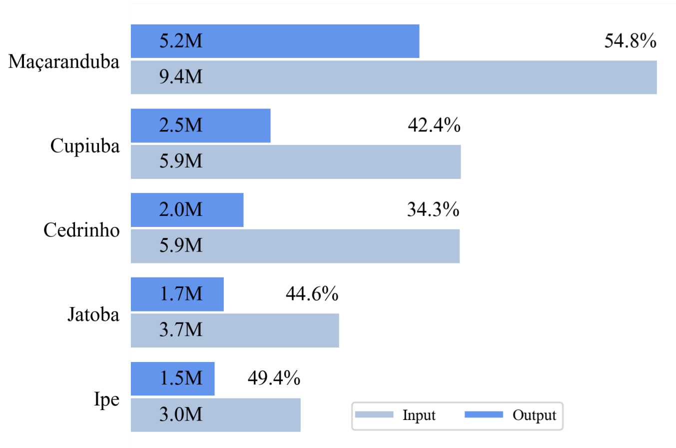
However, multiple supply chains may originate from a particular licensed forest v_f and converge at v. The volumes transported through these chains might exceed those of the supply chains connecting v to other licensed forests, therefore, they must be considered. We resort to Eppstein's algorithm[10] 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 the Dijkstra's second smaller and repeat the steps (2)-(4) considering now the second smaller Dijkstra path and so on until k chains are obtained. The proposed algorithm is detailed in the supplementary material.

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.

Results and Discussion

Analyzing 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) leaves the TTN through the end consumer nodes, which only receive wood from timber companies but do not send anything forward (sink nodes). The



	Nodes	Edges	Concessions
TTN - 578 Components	543860	779077	7450
Main Component	542587	778029	6559
TTN - 463 Components	467760	640010	6985
Main Component	466810	639222	6216
TTN - 308 Components	685668	860934	5387
Main Component	684808	860232	4652
TTN - 729 Components	155216	270640	8282
Main Component	153637	269379	7083
TTN - 809 Components	54970	95878	6720
Main Component	53001	94377	5427

Figure 1: 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. For each wood type, the first row shows the total number of nodes, edges, and licensed forests (source nodes from which timber enters the network) present in the TTN, while the second row provides the same information but is restricted to the largest connected component.

bar plots on the left in Figure 1 show the TTNs' input (light blue) and output (dark blue) volumes for five high-value hardwoods traded in Brazil: Ipe, Jatoba, Cedrinho, Cupiuba, 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 in percentage. As one can see, a substantial amount of wood vanishes, mainly due to losses inherent in the sawing process. Those balances get within the loss range pointed out in the literature[11] (35% to 54% waste depending on the wood type).

The table on the right in Figure 1 provides information about the TTN associated with each type of wood. For each wood type, the first row displays the overall count of nodes, edges, and licensed forests (source nodes from which timber enters the network) present in the TTN. The second row for each species provides the same information but is restricted to the largest connected component of the TTN. Notice that all TTNs have a main connected component containing about 99% of the nodes, except the Ipe, whose main component contains 96% of the nodes.

Figure 2 left shows the five largest connected components of the Ipe's TTN. Notice that

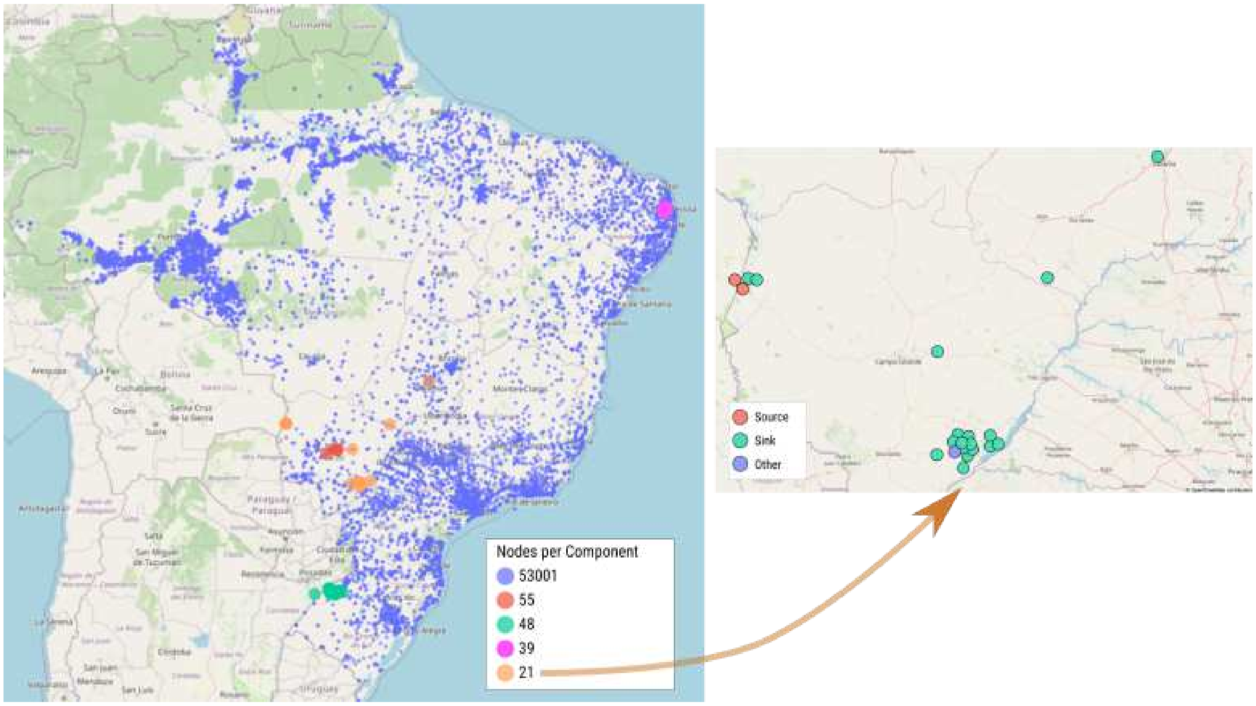
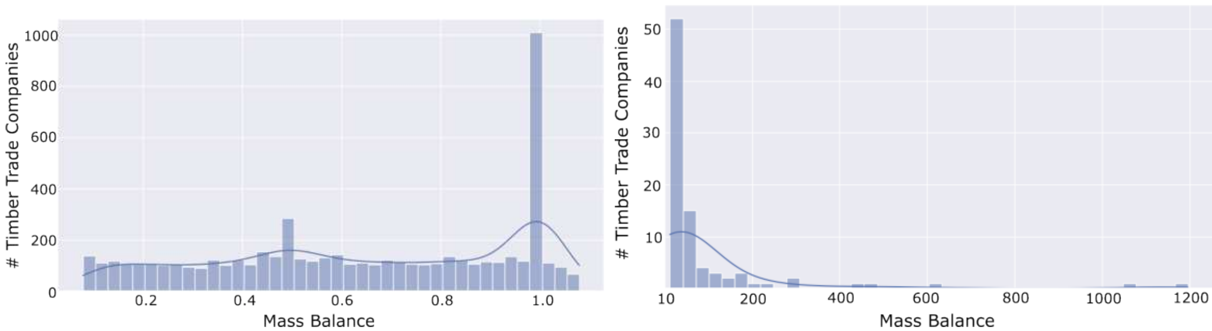


Figure 2: The five largest connected components for the Ipe's TTN.

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 behavior is also observed for the other wood species. Although the total volume of wood entering the fragmented components corresponds to less than 2% of the total volume entering the TTN, some of the fragmented components have unexpected conduct.

Figure 2 right depicts a zoomed view of the fragmented component with 21 nodes, which is located in the state of Mato Grosso do Sul. This fragmented component lacks a licensed forest node within it; specifically, the two source nodes (red circles in Figure 2 right) are not licensed forests, i.e., they are timber trade companies solely delivering timber, which raises suspicions. To enable this behavior, 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 potentially be sustained for a brief period if timber stored in the yards is used. However, the highlighted



(a) (b)

Figure 3: Mass balance of companies in the Ipe's TTN.

fragmented component has been operating for nine years, suggesting that timber is being input without proper registration by DOFs, which is illegal. We have also observed similar behavior in fragmented components involving other TTN species.

Odd Mass Balances and Supply Chains

As mentioned before, losses due to sawing processes are expected to happen. Therefore, the overall mass balance of timber companies, in terms of outflow $\times$ inflow, is expected to be less than or, in the best case, close to one. The histogram in Figure 3a shows that 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 in the TTN of 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 raw, 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.

However, as depicted in Figure 3b, some companies exhibit "spurious" mass balances, even exceeding $1,000$, indicating they have sent out much more timber units than they received. The substantial discrepancies in behavior certainly warrant attention, but they

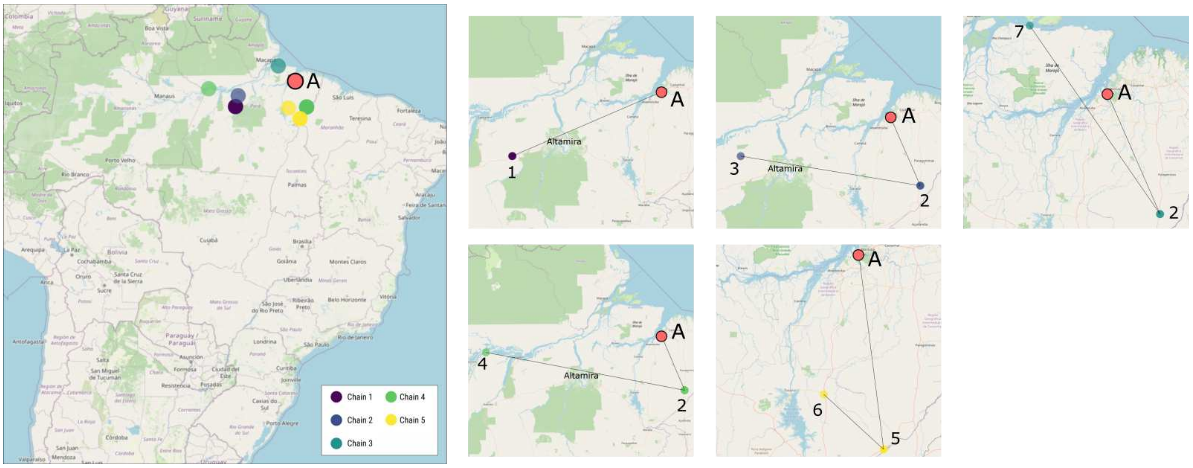


Figure 4: Top 5 most likely supply chain for company A (highlighted in red). The chains start in five different licensed forests.

may not necessarily indicate illegal conduct. For instance, the two companies with mass balances exceeding $1,000$ (rightmost in Figure 3b) were nearly inactive between 2010 to 2020, receiving approximately $2m^3$ of timber and dispatching around $2,000m^3$ during this period, a volume that could have been readily stored in their yards.

Nonetheless, there are enterprises whose behavior is not so simple to explain. As a case study, we examine a company with a mass balance of approximately 13, meaning that thirteen times more units of timber left the company compared to what entered it. For the sake of clarity, we will denote this company as A, and it is located in the state of Pará, as depicted by the red circle in Figure 4. Company A was fully operational between 2010 to 2020, receiving $3,665m^3$ of timber while forwarding approximately $46,530m^3$. Analyzing the time interval of inflow edges, we see that timber only input company A in 2011 and 2012, but wood products outflow from it all over the eleven years, the majority for export to other countries. Moreover, about 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 about 13, as the waste is considerable.

To further analyze company A, we computed its five most likely supply chains. The snapshots on the right in Figure 4 show the geolocation of the enterprises involved in the computed chains, which are numbered from 1 to 7 in the snapshots. Notice that the five most likely chains originate from five distinct licensed forests (the licensed forests correspond to the end of the chains not highlighted in red), which collectively input $18,762m^3$ of raw logs into the chains, significantly less than the $46,530m^3$ 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[1].

Upon searching the infraction notice database available on the Brazilian Institute of the Environment and Renewable Natural Resources (IBAMA)[2], 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 the submission of fake information into the control systems and the export of forest products without compliance with the mandatory "Regime Especial de Transporte" required for exporting timber products. Companies 2 and 5, which operate as intermediaries in four supply chains, had two and twenty-one registered 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 the market partners involved and its supply chains.

Discussion

The importance of timber traceability has leveraged some initiatives from governmental and non-governmental bodies in Brazil. For example, IBAMA released in 2020 the "DOF+Traceability",[11,12] 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[13] is another alternative explored by non-governmental agencies to match the unique composition of the timber's carbon, nitrogen, and oxygen atoms with those found in specific forest regions. Both DOF+Traceability and stable isotopes bear drawbacks that

[1] https://brasil.mapbiomas.org/en/2021/07/09/regiao-norte-lidera-desmatamento-no-brasil/

[2] `https://dadosabertos.ibama.gov.br/dataset/fiscalizacao-auto-de-infracao`

hamper their practical use, such as the possibility of fraud in the unique code generation and felling trees from areas nearby licensed forests with the same isotope signature, as illegal logging tends to occur around authorized areas.[14]

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 analyzing TNNs, consumers and government control agencies can identify groups of companies operating without connections to licensed forests, which serves as an indication that control systems are being circumvented. Moreover, investigating the most probable supply chains allows for the identification of the likely sources of the wood and the assessment of entities involved in these chains. As demonstrated in the presented case study, companies engaging in misconduct often form partnerships with other non-compliant enterprises, revealing trade networks that support illegal logging activities.

Therefore, the analytical capability provided 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 presented 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.

We are currently integrating the proposed methodology, specifically the supply chain analysis, into a web-based application that will be freely accessible to consumers and government enforcement agents.

Acknowledgement

The authors thanks Rodrigo Contreras, Dayse de Almeida, João Mattos, João Marcos Cardoso da Silva, and Fernando Nunes da Silva for their contribution in the early stages of data preparation and database construction. This work was partially supported by CNPq (grant #307184/2021-8) and Fapesp (grant #2022/09091-8). The opinions, hypotheses, and conclusions or recommendations expressed in this material are responsibility of the author(s) and do not necessarily reflect CNPq and Fapesp views.

References

(1) Barlow, J.; Lennox, G. D.; Ferreira, J.; Berenguer, E.; Lees, A. C.; Nally, R. M.; Thomson, J. R.; Ferraz, S. F. d. B.; Louzada, J.; Oliveira, V. H. F.; others Anthropogenic disturbance in tropical forests can double biodiversity loss from deforestation. *Nature* **2016**, *535*, 144–147.

(2) Nobre, C. A.; Sampaio, G.; Borma, L. S.; Castilla-Rubio, J. C.; Silva, J. S.; Cardoso, M. Land-use and climate change risks in the Amazon and the need of a novel sustainable development paradigm. *Proceedings of the National Academy of Sciences* **2016**, *113*, 10759–10768.

(3) Pearson, T. R.; Brown, S.; Murray, L.; Sidman, G. Greenhouse gas emissions from tropical forest degradation: an underestimated source. *Carbon balance and management* **2017**, *12*, 1–11.

(4) Ellwanger, J. H.; Kulmann-Leal, B.; Kaminski, V. L.; Valverde-Villegas, J.; Veiga, A. B. G.; Spilki, F. R.; Fearnside, P. M.; Caesar, L.; Giatti, L. L.; Wallau, G. L.; others Beyond diversity loss and climate change: Impacts of Amazon deforestation on infectious diseases and public health. *Anais da Academia Brasileira de Ciências* **2020**, *92*.

(5) Silva Junior, C. A. d.; Lima, M.; Teodoro, P. E.; Oliveira-Júnior, J. F. d.; Rossi, F. S.; Funatsu, B. M.; Butturi, W.; Lourençoni, T.; Kraeski, A.; Pelissari, T. D.; others Fires drive long-term environmental degradation in the Amazon Basin. *Remote Sensing* **2022**, *14*, 338.

(6) Martins, W. B. R.; de Matos Rodrigues, J. I.; de Oliveira, V. P.; Ribeiro, S. S.; dos Santos Barros, W.; Schwartz, G. Mining in the Amazon: Importance, impacts, and challenges to restore degraded ecosystems. Are we on the right way? *Ecological Engineering* **2022**, *174*, 106468.

(7) Lapola, D. M.; Pinho, P.; Barlow, J.; Aragão, L. E.; Berenguer, E.; Carmenta, R.; Liddy, H. M.; Seixas, H.; Silva, C. V.; Silva-Junior, C. H.; others The drivers and impacts of Amazon forest degradation. *Science* **2023**, *379*, eabp8622.

(8) CONAMA Resolution No. 379 19.10.2006 (Brazil, 2006). `https://shorturl.at/aimqZ`.

(9) Cormen, T. H.; Leiserson, C. E.; Rivest, R. L.; Stein, C. *Introduction to algorithms*; MIT press, 2022.

(10) Eppstein, D. Finding the k shortest paths. *SIAM Journal on computing* **1998**, *28*, 652–673.

(11) Franca, C. S.; Persson, U. M.; Carvalho, T.; Lentini, M. Quantifying timber illegality risk in the Brazilian forest frontier. *Nature Sustainability* **2023**, 1–11.

(12) CONAMA Resolution No. 497 10.09.2020 (Brazil, 2020). `https://shorturl.at/jKPSV`.

(13) Deklerck, V. Timber origin verification using mass spectrometry: challenges, opportunities, and way forward. *Forensic Science International: Animals and Environments* **2022**, 100057.

(14) Finer, M.; Jenkins, C. N.; Sky, M. A. B.; Pine, J. Logging concessions enable illegal logging crisis in the Peruvian Amazon. *Scientific reports* **2014**, *4*, 4719.

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.

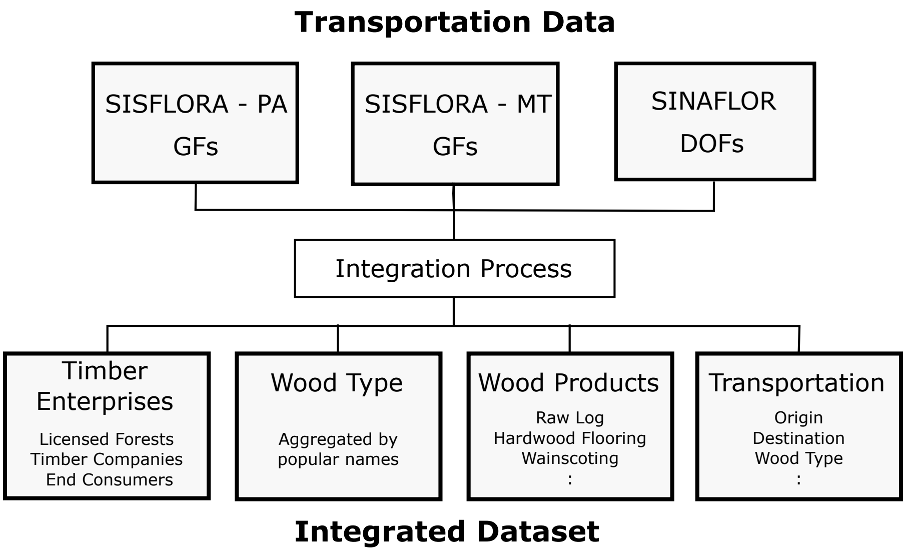


Figure 1: Data from the three control systems operating in Brazil, SISFLORA-PA, SISFLORA-MT, and SINAFLOR, are integrated into a single standardized dataset made up of four components: a table of unique enterprises (licensed forests, timber trade companies, and end consumers), a table with standardized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

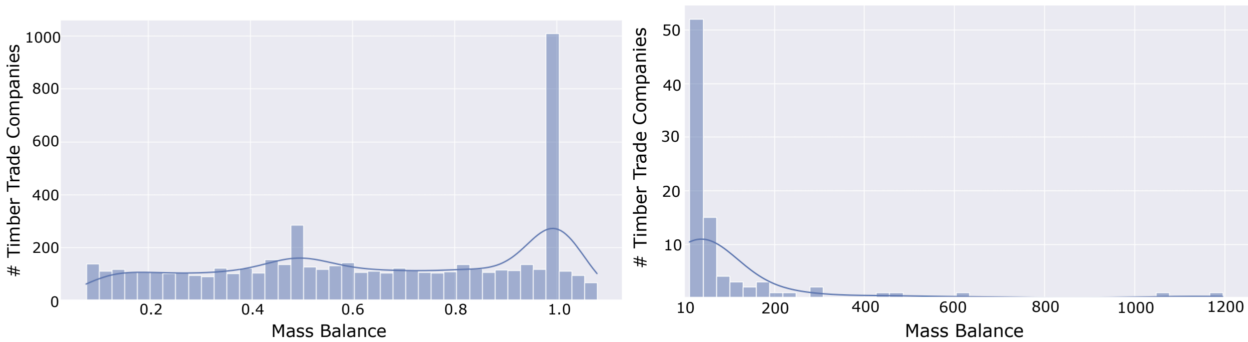
$count \leftarrow 0$
if $k > |C|$ **then**
  $k \leftarrow |C|$
end if
$i \leftarrow 0$
while $count < k$ **do**
  $\mathcal{C} \leftarrow \mathcal{C} \cup C[i]$ ▷ Add the likely chain $C[i]$ to $\mathcal{C}$
  $count \leftarrow count + 1$
  $i \leftarrow i + 1$
  $P \leftarrow$ Eppstein($G_{\mathcal{S}}$,C[i-1][0],C[i-1][-1],k) ▷ 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



(a) (b)

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.