

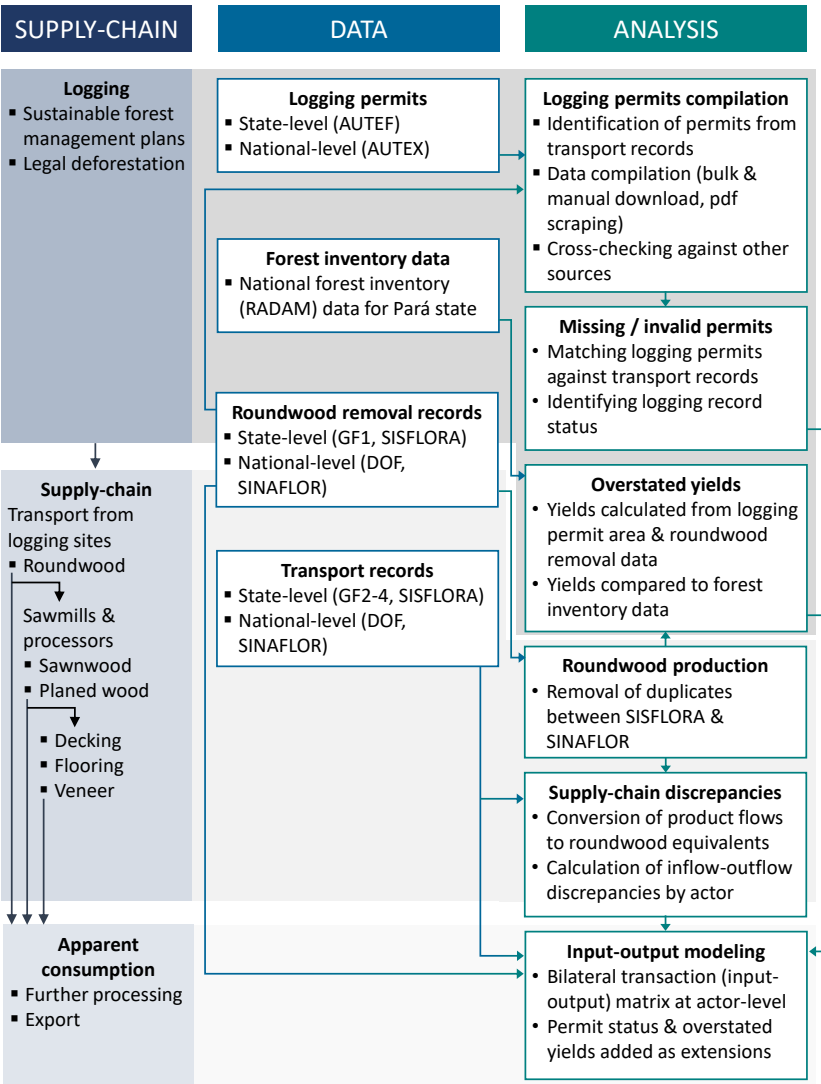
# Quantifying timber illegality risk in the Brazilian forest frontier

---

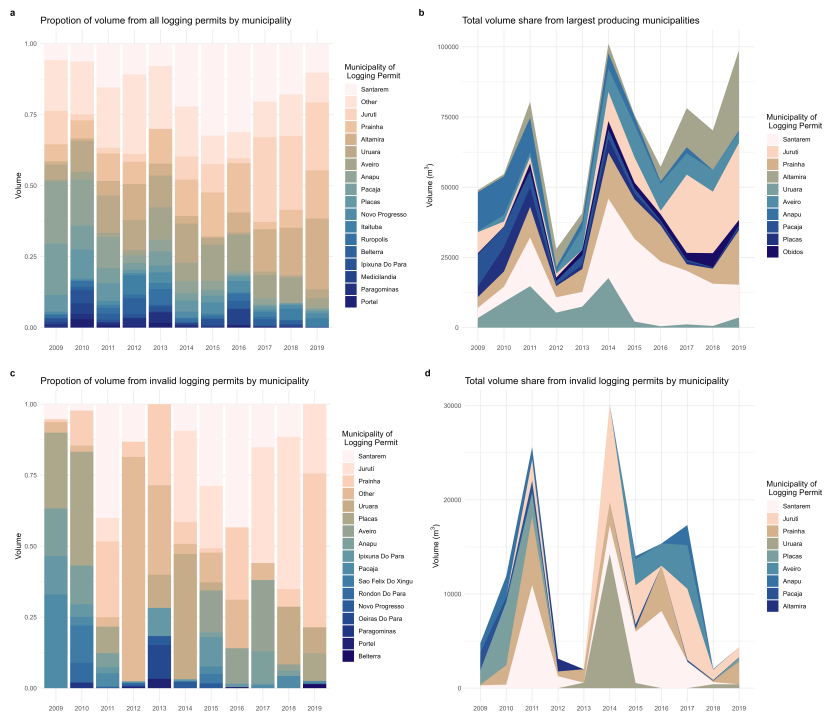
In the format provided by the  
authors and unedited

|                                                                                                 |          |
|-------------------------------------------------------------------------------------------------|----------|
| <b>Supplementary Figures</b>                                                                    | <b>2</b> |
| S1. Overview of methods . . . . .                                                               | 2        |
| S2. Temporal evolution of production by municipality . . . . .                                  | 3        |
| S3. Contrasting species yield between logging permits and RADAM<br>national inventory . . . . . | 4        |
| S4. Differentiated RADAM baseline according to area and yield<br>relationships . . . . .        | 5        |
| S5. Origin of overstated yields . . . . .                                                       | 6        |
| S6. Overview of SINAFLOR and SISFLORA-PA flows . . . . .                                        | 7        |
| S7. Yield, Volume and Area distribution by status of logging permit                             | 8        |
| <b>Supplementary Tables</b>                                                                     | <b>9</b> |
| ST1. Products and respective conversion factors . . . . .                                       | 9        |
| <b>Supplementary Notes</b>                                                                      | <b>9</b> |
| SN1. Logging permit compilation . . . . .                                                       | 9        |
| SN2. Estimating ipê yields and baseline stocking rates . . . . .                                | 11       |
| SN3. Rationale for flow inclusion . . . . .                                                     | 15       |
| SN4. Combining timber origin control systems . . . . .                                          | 18       |
| SN5. Software . . . . .                                                                         | 21       |

# Supplementary Figures

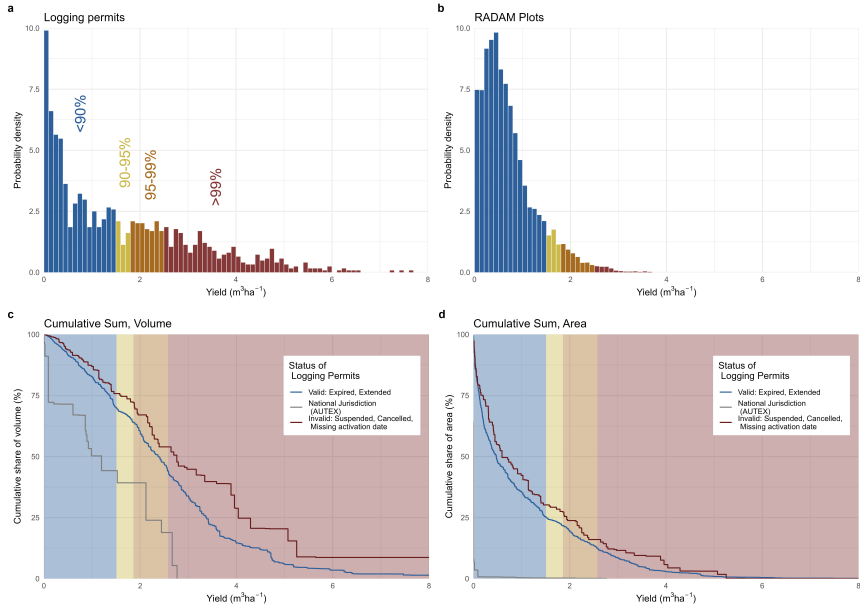


**Fig. S1** Overview of the different components of the analysis (right), what data is used in the different steps (center), and how this links to the different stages of the timber supply chain (left).

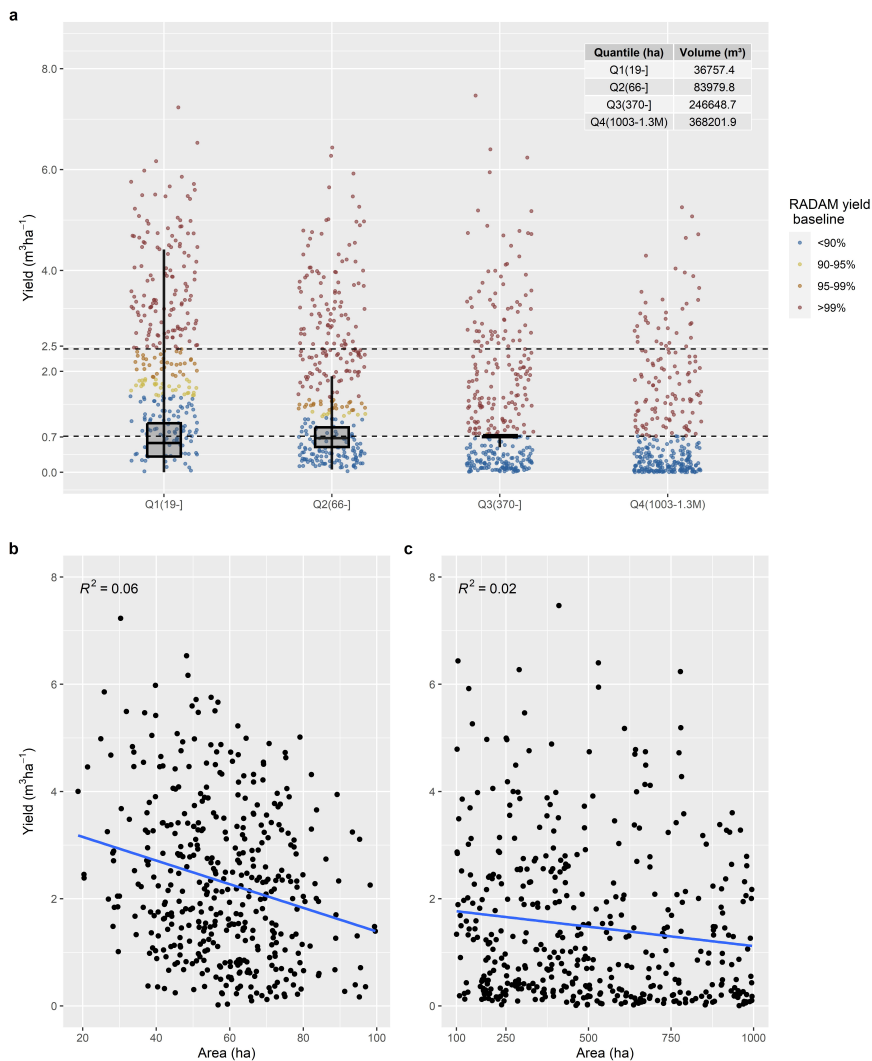


**Fig. S2** Temporal evolution of production by municipality for volume originating from all logging permits (a-b) versus only for invalid ones (c-d).

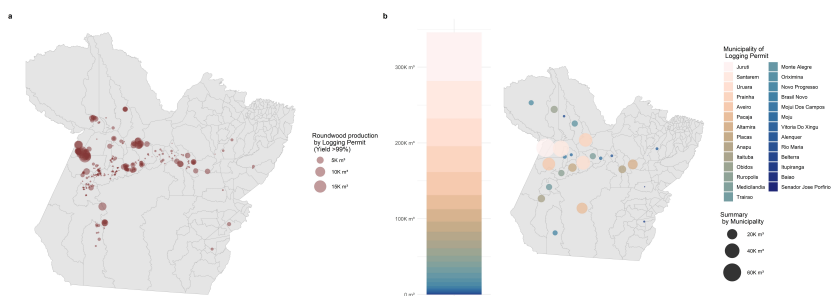




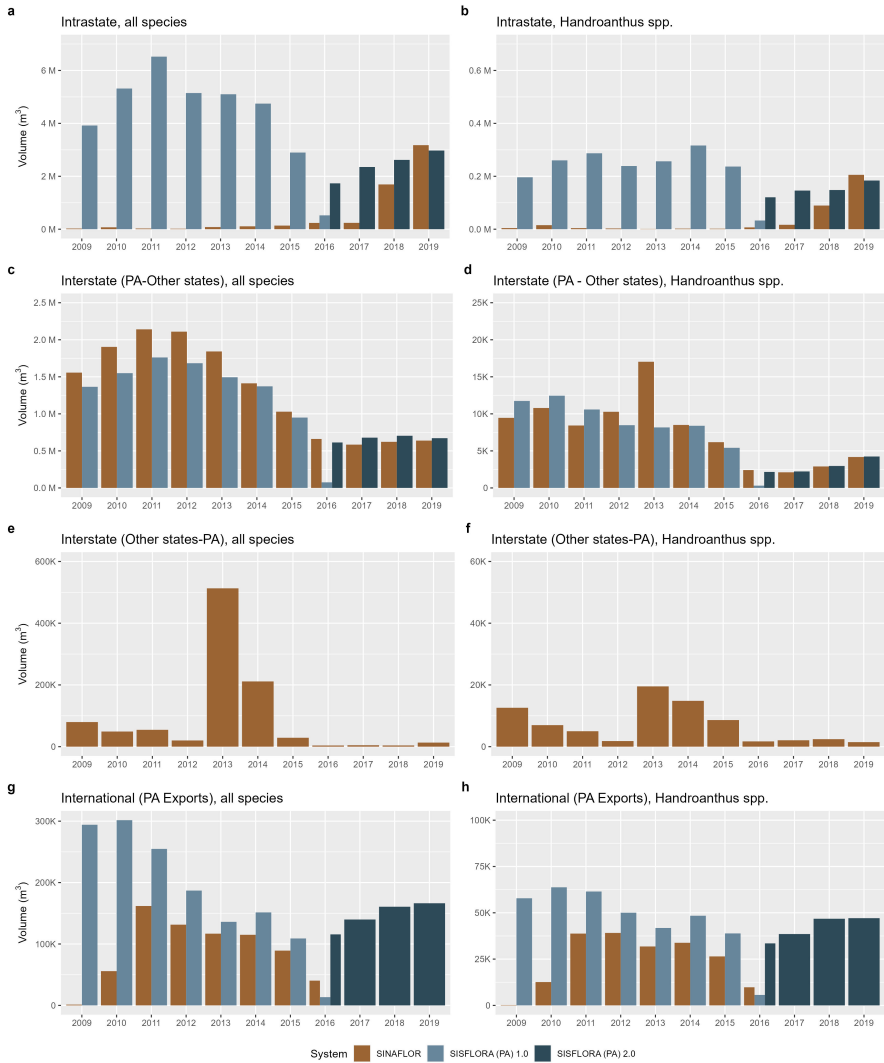
**Fig. S3** Contrasting species yield between volume entering the supply chain and RADAM national inventory. *Ipê* yield density for our data (a) in contrast to the national RADAM forest inventory (b). Colors show the intervals <90th, 90-95th, 95-99th and >99th percentiles in the RADAM inventory (all panels). Panels (c) and (d) show the cumulative share of volume and area, respectively, for logging permits of different status, highlighting that the large share of total volume (48%) that is >99th percentile yield comes from a relatively small area (14%).



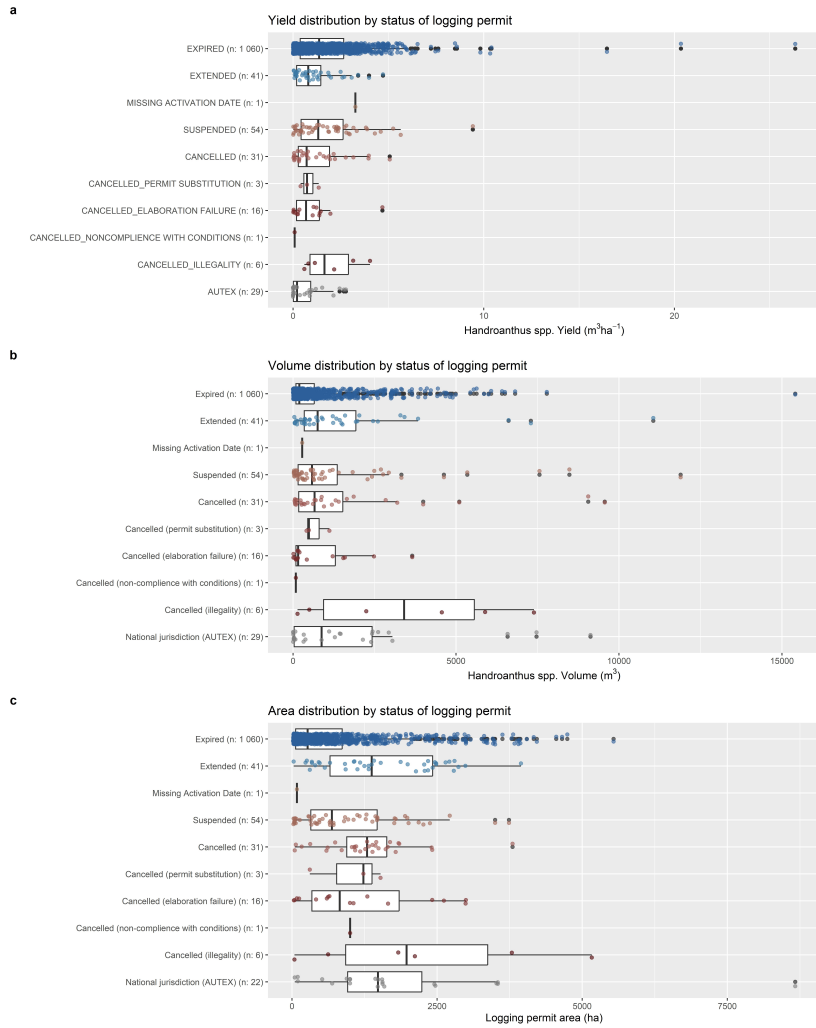
**Fig. S4** RADAM yield distribution simulation according to the lowest value of each area quartile (19, 66 and 370 ha) (a). Summary volumes in (a) refers only to valid (expired and extended) permits. The black box contains the estimated 25th to 75th percentiles, with the center line denoting the median (50th percentile). The end of the lines mark the 0th and 100th percentiles. Observations are based on simulation ( $n = 10,000$ ). Relationship between areas under 100ha and yield ( $n = 419$ ) (b) and between 100>1000ha and yield ( $n = 438$ ) (c). The top ten largest yields were removed from all figures.



**Fig. S5** Location of logging permits with yield above the 99th percentile class (a). Volume from overstated permits entering the supply chain summarized by municipality of origin (b).



**Fig. S6** Summary of total flows in the three different reporting systems: SINAFLO (national-level), SISFLORA (Pará state-level system) 1.0 (pre-2016) and 2.0 (post-2016). Panels a, c, e, g show an overview of transacted timber products for all species, and panels b, d, f, h for *Handroanthus* spp. (and *Tabebuia* spp. now *Handroanthus*, see methods) alone, at distinct jurisdictional levels. No conversion has been applied at this stage.



**Fig. S7** Yield (a), volume (b) and area (c) distributions by status of logging permits. The black box contains the 25th to 75th percentiles, with the center line denoting the median (50th percentile). The end of the lines mark the 0th and 100th percentiles, and values beyond these upper and lower bounds are considered outliers. Area distributions in (c) does not display seven logging permits with areas above 10,000 ha, all under the "National Jurisdiction (AUTEX)" category and included in the overall analysis.

# Supplementary Tables

**Table 1** Products and respective conversion factors

| Products (EN)                                                                                                                                                                                | Products (PT)                                                                                                                                                                                            | Low; Mid; High (VYC <sup>1</sup> %) |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------|
| Primary breakdown: block, square of fillet, sawnwood (includes board, beam, small beam, rafter, short rafter, plank, wide plank, batten, short batten, clapboard, short clapboard, crosstie) | Processamento primário: bloco, quadrado ou filé, madeira serrada (incluindo tábua, viga, vigota, caibro, caibro curto, prancha, pranchão desdobrado, sarrafo, sarrafo curto, ripa, ripa curta, dormente) | 35; 44.5; 53.9                      |
| Planed 2 faces (S2S)                                                                                                                                                                         | Madeira aplainada 2 faces (s2s)                                                                                                                                                                          | 29.8; 37.8; 45.8                    |
| Planed 4 faces (S4S)                                                                                                                                                                         | Madeira aplainada 4 faces (s4s)                                                                                                                                                                          | 28.7; 36.5; 44.2                    |
| Decking, flooring                                                                                                                                                                            | Decking, pisos e assoalhos                                                                                                                                                                               | 28.7; 36.5; 44.2                    |
| Sliced veneer                                                                                                                                                                                | Lâmina faqueada                                                                                                                                                                                          | 45 <sup>2</sup>                     |
| Peeled veneer                                                                                                                                                                                | Lâmina torneada                                                                                                                                                                                          | 55 <sup>2</sup>                     |

Source: Application of general norms CONAMA 474 2016[54] and CONAMA 497 2020[58] and species-specific figures Romero et al.[57]

<sup>1</sup>Volumetric Yield Coefficient

<sup>2</sup>Sliced and peeled veneer remained the same given marginal share and lack of species-specific baseline figures

## Supplementary Notes

### SN1. Logging permit compilation

The lack of a comprehensive database of logging permits that detail the origins of forest products—despite recent efforts (<https://www.gov.br/ibama/p/t-br/assuntos/biodiversidade/flora-e-madeira/paineis-analiticos-da-gestao-madeireira-1>) and a longstanding mandate for environmental authorities to make data available to public [59]—have precluded an assessment of the extent to which invalid logging permits are used to substantiate credits in the along the timber supply chain. To overcome this, we compiled a consolidated logging permits dataset for the state of Pará by combining different data sources made

available publicly (via official government webpages) or through other publications. For permits issued to enterprises under national jurisdiction, details are made available via the Brazilian Open Data Portal via the "authorization" (*autorizacao*) module of the DOF dataset <https://dados.gov.br/dados/conjuntos-dados/dof-autorizacoes-de-exploracao-florestal>, including coordinates of origin and authorized area, authorization type and status, species and volumes.

For permits issued under the state jurisdiction—accounting for the bulk of production volumes in the state of Pará—this information is primarily made available via PDFs listed under the search engine at the SIMLAM (*Sistema Integrado de Monitoramento e Licenciamento Ambiental*) webpage of the Pará Environmental Secretariat (SEMAS, <https://monitoramento.semas.pa.gov.br/simlam/index.htm>). The webpage recently (March, 2023) underwent a major update, with webpage link now redirecting to the newly launched "Transparency Portal" <http://portaldatransparencia.semas.pa.gov.br/>, where logging permit information can now be found. A georeferenced polygons file was also accessed via the former webpage, but it covered only 54% of *ipê* volume. In order to complement and validate the latter, we downloaded all PDFs listed in the SIMLAM webpage ( $n = 2,259$ ), matching those to permit numbers listed in the GF and DOF records. While 887 of these permits were used to back up *ipê* flows, this revealed that a number of logging permits ( $n = 365$ ) were still missing. We manually entered the missing permits numbers into the SIMLAM search engine, finding that these logging permits in most cases ( $n = 328$ ) could be located in case a specific number was searched, despite not being displayed in the list of available logging permits.

This resulted in a dataset of 2,571 state-level and 193 federal-level logging-permits, totalling of 2,764 unique permits with comprehensive data. From the

transport side, 1,292 permits have been used by supply chain actors transporting roundwood to report place of *ipê* origin: 1,215 of these were backed up by logging permits issued at state-level, 47 at federal-level, while the remaining 30 could not be backed up (original permit data could not be downloaded or found). We cross-checked our compiled logging permit database against the data in Brancalion et al. [18], IBAMA (2019) [30], as well as original permits listed on the DOF "authorization" module for any additional information.

As highlighted in the Results, nearly all *ipê* production comes from Sustainable Forest Management Plans (PMFS), most of which (92% of volume) have logging permits authorized at state-level and 7% originated from enterprises licensed under national jurisdiction, mainly concessions and—to a lesser extent—community-based forest management operations. The share under national jurisdiction has been increasing over time, though, from only 2% of logging permits in 2009-2015, up to 25% of logging permits by 2019, following the start of extraction from the National Forest concessions of Saracá-Taquera (2013), Altamira (2016) and Caxiuanã (2018) [60], and it can be expected to increase with planned expansion of granted concessions [60]. It is worth noting we found only a insignificant amount (0.1%) comes from legal deforestation (or as referred to across regulations, vegetation suppression).

## SN2. Estimating *ipê* yields and baseline stocking rates

To calculate implicit (stated) *ipê* yields for the state-level permits (AUTEFs), the area used was as indicated by the effective area (*Área Líquida*, ha) of the Annual Production Unit (*Unidade de Produção Annual*, UPA). For the national level permits (AUTEX) data, we used the only area field listed in the DOF "authorization" module. From the 1,292 permits listed as supplying *ipê*, 1,243 could be used to derive yield figures. The remaining 49 either had areas



set to zero (16 permits from AUTEX, all under type *supressão da vegetação* or "vegetation suppression", the category for legal deforestation) or missing areas altogether (in cases where permits could not be accessed or found). We investigated outliers both in terms of areas and yields in order to detect any error in data processing, but found all corresponded to reported data and thus did not remove such values from the analysis. Fig. S7 provides a visual summary of volumes, areas and yields included.

A few assumptions and limitations are relevant clarifying when reconciling logging permits and transport records with RADAM data. On species inclusion, we opted for all known and commercialized as *ipê*, making use of all records that listed *Handroanthus* spp. or *Tabebuia* spp. at genus or species level. The RADAM plots available provided information on two *ipê* species—*T. impetiginosa* and *T. serratifolia*—reflecting the fact RADAM data was collected in the early 70's when the *Handroanthus* genus was just being proposed and a number of other species of the *Tabebuia* spp. genus were yet to be described[61]. Still, these two species make up 87% of all roundwood being reported entering the supply chain for the period under analysis. An additional 8% has been reported at *Tabebuia* sp. genus-level only, with the remaining 5% of volume traced to other 9 reported species. In 2019 (latest year of analysis) *ipê* species share were 79% *H. serratifolius*, 21% *H. impetiginosus*, 0.5% *T. capitata*, 0.3% *H. incanus*, 0.004% *T. chrysantha* (with no reporting at genus level), which only occurred in this year at a marginal share for decking and sawnwood categories. Indeed, it was only in 2007 after a major revision of the *Tabebuia* sp. genus that *Handroanthus serratifolius* (Vahl) S. Grose, *Handroanthus impetiginosus* (Mart. ex DC.) Mattos, and other *ipê* species started to be distinguished. As of 2008, *Tabebuia* have been generally not distinguished commercially by species, with *ipê* species often being lumped together (including

in field assessments) [31]. Thus, based on considerations of RADAM historical data collection context, the species revisions since then, patterns of reporting where species were listed across time at genus level or reported broadly as *ipê* as well as the broader aim of this study, we find appropriate to use all species under the two genera, grouping which makes yield figures from reported volumes entering the supply chain comparable to RADAM estimates.

On *ipê* distribution and abundance, Schulze et al.[31] aimed to tentatively demarcate regions of relatively high, medium and low mean densities across the Amazon and thus it raises the question on whether a more representative figure for sub-regions within Pará would be preferable. Firstly, both number and size of plots have been pointed as limiting factors for the representatives of RADAM figures reported by Schulze and Brancalion, but they continue to be seen as a key dataset for deriving estimates of populations distribution [50, 62]. While larger and more recent inventories may provide a more representative distribution in state sub-regions, RADAM is still the sole Brazilian Amazon-wide inventory with data collected prior to the expansion of the deforestation and forest degradation frontier. Thus, even after data on the ongoing country-wide inventory is made available for all Amazon states, RADAM will remain an important source of information to derive population densities baselines at a time of minimal development in the region.

Yet, it raises the question on its limitations for the elaboration of species distributions at a more granular scale. The assumption areas of PMFS are expected to display average yields of high-value species neglects the selection bias of logging towards areas with high densities of such species. The very occurrence of these is a key determinant on whether timber exploitation is economically viable in the first place [22, 31] and thus it can be expected yields in management areas to be above average. We have opted for the use

of yield above the 99th percentile of the 19 ha-based distribution to convey the constraints of such selection bias. Different approaches can be tested and thresholds set to illustrate what differentiated risk figures would entail. Fig. S4 panel (a), for example, demonstrates the implications of constructing an yield distribution with increasingly large areas: the likelihood these may exceed a given 99th percentile threshold becomes increasingly small, with significantly more risk being assigned to volumes originating from relatively larger areas in the case a differentiated approach to risk would be undertaken. Likewise, Fig. S4 panels (b) and (c) show the yield distribution according to smaller logging permit area and highlight that—despite certain bias could be observed particularly in small areas—yield distribution show no clear trend, with relatively low as well as relatively high yield occurring in the first three area quartiles. Furthermore, disregarding risk in the first area quartile (removing risk from all areas between 19-66ha), decreases the illegality risk from overstated risk from 48.1% of total valid permit volume to 44.5%, which contrasts with adding to this baseline volumes  $> 0.7m^3/ha$  for the largest quartile alone (73.2%).

It is also worth pointing out that access to location and other transportation infrastructure, access to processing hubs and markets, the legitimization of forestland access and ownership, as well as actors' operations in these geographies are also factors that influence forest operations and may lead to the exploitation of areas with sub-optimal (or average) species distributions. Moreover, despite uncertainties of what then should constitute an non-biased baseline species distribution, Brancalion et al. [18] points to the fact RADAM estimates are conservative, given that these do not remove the 10% species retention legally required in a PMFS (or applies the three individuals/100ha retention). They also find subtle techniques for overestimating yields, such as declaring DBH measurements larger than possible given stumps in the

field. The figure also does not account for sub-merchantable individuals, which according to inventory data used in Schulze et al[31] are anywhere between 12–37% of individuals ”with an additional unknown fraction possessing cryptic hollows” (p.2075). In the end, such observations point to a conservative figure in what relates to the threshold used for determining yield overestimation, being at the same time comparable and replicable across key species.

### **SN3. Rationale for flow inclusion**

#### **Production vector**

For the SISFLORA-PA system, roundwood transported from forest of origin is registered under GF1, a category specifically used for this purpose [63]. While the state-level system encompasses the majority of production (Fig. S6), careful attention is warranted in accounting for flows that may be displayed in SINAFLORE and not SISFLORA-PA, given flows from national jurisdiction enterprises (e.g. concessions, certain community-based forest management) may be accounted for in the former, but not the latter. Additionally, roundwood re-transport is also treated differently between the state- and national-level systems and, hence, we looked to minimize any double-counting arising from re-sale and subsequent re-transport, as well as from reporting duplication between the national and state-level systems.

Starting with SISFLORA-PA, production volume overestimation coming from re-transport was addressed by the removal of all roundwood flows reported under GF3, GF3i and GF4, as these categories are only used for roundwood in the case of re-sale and re-transport [63]. It is worth emphasizing such removals, because the remaining subset not only constitutes a better approximation of true production figures, but also prevents the overestimation of illegality risk from missing permits, since GF3, GF3i and GF4 categories do

not list logging permits of origin (as should all GF1-level transactions). Within the GF1-level subset, we removed flows where actors sent roundwood to themselves, given that such flows likely represent some form of re-transport (and as such have no implication on the physical balance of the input-output analysis). The latter removals accounted for about 11% (75,672 m<sup>3</sup>, 5,970 records) in SISFLORA 1.0 and 6% (19,843 m<sup>3</sup>, 7,531 records) in SISFLORA 2.0.

Under SINAFLOR, no similar categorization of DOF types is available. According to IBAMA [30], re-transport of roundwood can be broadly removed by discarding flows which states "*pátio*" (yard) as the type of origin. Nonetheless, we found that only about 35% of all roundwood flows for *ipê* with origin in Pará state had a stated type of origin. This implies a different approach was needed to avoid inflating production figures: We start by discarding all flows with no reported logging permit (150,610 m<sup>3</sup>, 66% of SINAFLOR roundwood flows), given the obligation on reporting location when roundwood first leaves the forest of origin. From the remaining subset, we removed all flows that had yard listed in their type of origin (2,444 m<sup>3</sup>, 3%). Consistent with the approach adopted at state-level, we also removed flows where origin was equal to destination (2,451 m<sup>3</sup>, 3%). Lastly, to avoid double-counting, we searched for any flows where actor and municipality of origin summed to equal volume between the two systems, but found no instances of this. Indeed, we found that the bulk of removals (148,236 m<sup>3</sup>, 64% of total SINAFLOR roundwood flows) had four key features in common: they (i) did not have a listed permit, (ii) had actor of origin that equaled that of destination, (iii) had data on type of origin missing, and (iv) did not display coordinates of origin or destination. It is likely the flows that share this lack of reporting altogether encompass duplication in reporting between systems and hence likely the reason why no further duplicates were found.

## Bilateral transaction matrix

Regarding product selection, we aimed to include all major timber products relevant for *ipê*. Nonetheless removals are worth clarifying. From the outset, we exclude all categories that relate to Non-Timber Forest Products (e.g. foliage, bark, etc.), as well as all categories that refer to biomass for energy production (including firewood and charcoal) and residues of the processing industry in general. Notably, we exclude "finished products" from the analysis, mainly since the broadness of the category makes the selection of a representative conversion factor a challenge (if possible altogether). Additionally, this category is not widely used. IBAMA [30] shows it makes up 2.45% of the total final consumption for all species produced in Brazil between 2012-2017, but it does not make the top 10 products (97.55% of final consumption) for the case of *ipê*. We also excluded "short logs" (*toretes*)—mostly reported under GF2 in SISFLORA as a category for biomass, residues and other unprocessed products and by-products removed from the forest in general—due to the ambiguity of whether this is processed into other timber products or rather used as biomass/charcoal. Even if IBAMA [30] included this in their analysis, we find that this category accounted for only about 0.2 % (SISFLORA PA 1.0 and 2.0) and 0.6% (SINAFLOR) of the roundwood volume (with no removal of re-transport). Lastly, other products removed due to marginal flows over the entire time period are: "roundwood of produced timber" used for commercial plantations (42 m<sup>3</sup>), "posts" (47.2 st), and plywood (4.58 m<sup>3</sup>).

We also looked to minimize instances of double-counting between state and national systems. Given SINAFLOR data appears to have CPF/CNPJ of origin set equal to CPF/CNPJ of destination for instances of duplication between systems, and that such records carry no practical implication on the physical input-output balance, we discard all such flows upfront. However,

as setting CPF/CNPJ of origin to that of destination is also a feature of all export records in SISFLORA 1.0 and SINAFLOR, prior to removing these instances we ensure export flows are not removed. We also ensure we do not remove flows which are not exports, but known to be final consumption (i.e., in SISFLORA, when cadaster of destination is missing but CPF/CNPJ of origin equal destination), though these are minimal for *ipê*. On deduplication for the export share, given SINAFLOR is missing export data for Pará since 2016 and that these export figures are systematically reported below state-level value (Fig. S6), we opt to use export figures coming from SISFLORA-PA only, deemed appropriate given Pará’s focus in this study. Approximately 10% of bilateral transactions volume (362,605 m<sup>3</sup> or 8% 25,744 records before conversion) are discarded from removing origin that equal destination. Approx. 38% (265,006 m<sup>3</sup>) are from SINAFLOR, while 3.5% for SISFLORA 1.0 and 2.6% for SISFLORA 2.0. We discarded a further 3k m<sup>3</sup> from SINAFLOR by checking for equal actor-municipality sums between systems (Suppl. Text ST4 below for considerations on deduplication).

All flows were converted to roundwood equivalents, given that volumetric conversion factors are used to regulate allowable conversion rates (as well as form the basis for reporting by actors and monitoring by authorities) as mentioned in Methods. Nonetheless, other equivalent conversions may be considered. Carbon equivalents can be preferable to avoid double-counting when several products with different properties (e.g. sawnwood, paper, particle board, biomass, charcoal) are being considered [53]. However, in the case of *ipê*—aside from the broader use of RWE—decking, flooring and sawnwood constitute an overwhelming part of total consumption, with residues for energy making up only 0.5% of total final consumption [30].

## SN4. Combining timber origin control systems

Constructing a comprehensive and consistent dataset on timber transactions for the state of Pará requires cleaning and harmonizing data from three different datasets: the state-level (SISFLORA-PA) system prior to and after 2016 (versions 1.0 and 2.0, respectively), as well as the national-level (SINAFLOR) system. The upgrade of the SISFLORA-PA system in 2016 primarily allowed for the inclusion of more detailed stump and log tracing via RFID-based chip and bar code system, but also other system refinements. This has resulted in two slightly differing data structures, with the inclusion of new variables (e.g. log bar code, type of destination) on the one hand, and suppression (e.g. type of origin) on the other. Such differences are, however, minor when comparing to those between SISFLORA-PA and SINAFLOR, which includes not only differences in data structure, but also implications related to the jurisdiction where norms and procedures apply. Thus aside from data structure, the overlapping policyscape translates into reporting duplication between systems which has so far not been properly addressed by SINAFLOR.

In attempting to remove duplicate data entries (i.e., transport records that have been reported in both the SINAFLOR and SISFLORA systems), we found no optimal solution based on a set of unique identifiers drawn from the data collected and made available to us. Aside from the lack of a unique global key [64]—which could plausibly be expected from SINAFLOR given their institutional mandate of integration between systems—data on CPFs for actors of origin and destination have been partially anonymized in the SINAFLOR data, in line with recent national data protection law (LPGD) [65]. It is worth clarifying these differ from CNPJ, which are made openly available (e.g. for a data sample: <https://dadosabertos.ibama.gov.br/dados/DOF/PA/transporte/2021.html>). While there are benefits to individual’s data protection



(which can be done with methods that provide individual anonymization while still maintaining information of timber origin and commercialization), given the remaining variables that are present in both datasets, the CPF is important in supporting identification in the uniqueness of a given transaction throughout the period of analysis. This is particularly true moving from intrastate-only to national-level control: where there are needs for alignments of state-level cadasters with environmental authorities (e.g. CEPROF, *Cadastro de Exploradores e Consumidores de Produtos Florestais* or "Cadaster of Forest Products Exploiters and Consumers") to those of national-level cadasters (e.g. CTF, *Cadastro Técnico Federal* or "Federal Technical Cadaster").

In this study, we used the partial anonymization implemented in the SINAFLOR-level data (i.e. making available only the first three digits of CPF and only the initial of names associated with CPFs) in order to establish actor's uniqueness between SINAFLOR and SISFLORA-PA systems. First, while partially anonymized in SINAFLOR, details added to other variables (e.g. yard of origin and destination) still in SINAFLOR could be used to determine full CPF and individual identity (here a weakness of the anonymization method currently used by SINAFLOR). Last, by combining the first three digits of CPF and initials, actor uniqueness can still be validated with SISFLORA-PA data (as of March 2023—albeit incomplete—GF data has been made available through the newly available Transparency Portal from SIMLAM-SEMAS (<http://portaldatransparencia.semas.pa.gov.br/>), including GF original PDFs files and CPF data) Moreover, while we make no sensitive data available, individual's cadaster (i.e. CPFs) are an important key for deduplication given the lack of proper data harmonization from the part of government authorities, and more specifically SINAFLOR given its mandate.

Aside from the lack of global unique keys for removing duplicates between systems, further challenges can be better illustrated when systematic grouping multiple variables. Firstly, only testing the amount of roundwood for *ipê* (*Handroanthus* spp.) transported in 2019 (a year that broadly appears to have higher level of equal reporting between systems and thus higher duplication level, see Fig. S7b for example), SINAFLOR reports a total of 138,455 m<sup>3</sup> of which 79% state an origin that is equal to destination. When looking at the same parameters for SISFLORA-PA, volume totals 114,671 m<sup>3</sup> (a lower value can be expected, as SINAFLOR may encompass reporting for licensed enterprises at both state- and national-level), but with only 7% of actor of origin being the same as destination. Such large differences in reporting key information as origin and destination suggests intrinsic differences between datasets. Indeed, most SINAFLOR records for this year displayed four key features in common: they (i) did not have a listed permit (needed for transport from forest of origin), (ii) had actor of origin that equaled that of destination, (iii) had data on type of origin missing, and (iv) did not display coordinates of origin or destination. Thus, it is likely the flows that displayed this shared lack of reporting are the duplicates from state-level system. Given flows with origin that equal destination have no implication on the physical balance of the input-output analysis (see Methods section as well), we removed these, which by default minimize the potential double-counting. After the removal of records which displayed common features of duplicates (i.e. systematic lack of many data fields in SINAFLOR) we ensured no clear duplicates between systems remained. We summed up volumes per actor, municipality, product, specie and year and removed from SINAFLOR all instances where the sums were equal between the two systems.

## SN5. Software

All analyses pertaining to this study have been performed in R [66] and, to a lesser extent, Python and used several open source libraries. Notably scraping tools for the SEMAS-SIMLAM webpage as well as PDFs were built with Python (last replicated in Python 3.10 and 3.11 respectively). Administrative boundaries were obtained via API from geobr [67] R library. All steps that pertain to data cleaning, integration, manipulation and analyses as well as output visualization and mapping were performed in R (last replicated in R 4.3.1). Data cleaning, wrangling and analyses used particularly the tidyverse [68], janitor [69], sf [70], base R, but also tmap [71], expm [72], ggpmisc [73], reshape2 [74]. Graphs displayed throughout the manuscripts were based particularly on ggplot2 [75], scatterpie [76] and ggalluvial [77], but also on patchwork [78], scales [79], palettee [80] and scico [81] libraries. Other libraries used are referenced throughout the code made available for reproduction of this study available at <https://github.com/carolsrto/illegality-risk-ns>.

## References

- [1] Grantham, H. S. *et al.* Anthropogenic modification of forests means only 40% of remaining forests have high ecosystem integrity. *Nature Communications* **11** (1), 5978 (2020). <https://doi.org/10.1038/s41467-020-19493-3>.
- [2] Potapov, P. *et al.* The last frontiers of wilderness: Tracking loss of intact forest landscapes from 2000 to 2013. *Science Advances* **3** (1) (2017). <https://doi.org/10.1126/sciadv.1600821>.
- [3] Matricardi, E. A. T. *et al.* Long-term forest degradation surpasses deforestation in the Brazilian Amazon. *Science* **369** (6509), 1378–1382 (2020).

<https://doi.org/10.1126/science.abb3021>.

- [4] Barlow, J. *et al.* Anthropogenic disturbance in tropical forests can double biodiversity loss from deforestation. *Nature* **535** (7610), 144–147 (2016). <https://doi.org/10.1038/nature18326>.
- [5] Qin, Y. *et al.* Carbon loss from forest degradation exceeds that from deforestation in the Brazilian Amazon. *Nature Climate Change* **11** (5), 442–448 (2021). <https://doi.org/10.1038/s41558-021-01026-5>.
- [6] Pearson, T. R. H., Brown, S., Murray, L. & Sidman, G. Greenhouse gas emissions from tropical forest degradation: an underestimated source. *Carbon Balance and Management* **12** (1), 3 (2017). <https://doi.org/10.1186/s13021-017-0072-2>.
- [7] Silva Junior, C. H. L. *et al.* Amazonian forest degradation must be incorporated into the cop26 agenda. *Nature Geoscience* **14** (9), 634–635 (2021). <https://doi.org/10.1038/s41561-021-00823-z>.
- [8] Kleinschmit, D., Mansourian, S., Wildburger, C. & Purrett, A. *Illegal logging and related timber trade-dimensions, drivers, impacts and responses. A global scientific rapid response assessment report* Vol. 35 (IUFRO (International Union of Forestry Research Organizations) Secretariat, 2016).
- [9] FAO. Terms and definitions: FRA 2020. (2018). Available at: <http://www.fao.org/3/I8661EN/i8661en.pdf>.
- [10] Brazil. Ordinance N<sup>o</sup> 253 18.08.2006 (2006). Available at: <http://www.ibama.gov.br/sophia/cnia/legislacao/MMA/PT0253-180806.PDF>.

- [11] Brazil. CONAMA Resolution Nº 379 19.10.2006 (2006). Available at: <http://www.ibama.gov.br/sophia/cnia/legislacao/MMA/RE0379-191006.PDF>.
- [12] Brazil. IBAMA Normative Instruction Nº 21.12.2014 (2014). Available at: <http://www.ibama.gov.br/phocadownload/sinaflor/2018/2018-06-13-Ibama-IN-IBAMA-21-24-12-2014-SINAFLO-DOF-compilada.pdf>.
- [13] Perazzoni, F., Bacelar-Nicolau, P. & Painho, M. Geointelligence against illegal deforestation and timber laundering in the brazilian amazon. *ISPRS International Journal of Geo-Information* **9** (6), 398 (2020). <https://doi.org/https://doi.org/10.3390/ijgi9060398>.
- [14] MPF. Civil Lawsuit: Operation Arquimedes (Initial Petition, Process: 1002687-04.2019.4.01.3200), Federal Prosecution Office (MPF) (2019). Available at: <http://www.mpf.mp.br/grandes-casos/operacao-arquimedes/atuacao-do-mpf/acoes-civis-publicas/acoes-civis-publicas-operacao-arquimedes>.
- [15] Dittmar, H. Archimedes operation and the destruction of the Brazilian Amazon forests with the state consent. *RECIMA21 - Revista Científica Multidisciplinar - ISSN 2675-6218* **3** (3), e331260 (2022). <https://doi.org/10.47820/recima21.v3i3.1260>.
- [16] de Abreu, A., Toledo, L. F. & Goulart, E. How endangered brazilian timber ends up in the us (2022). Available at: <https://www.occrp.org/en/investigations/how-endangered-brazilian-timber-ends-up-in-the-us>.
- [17] Greenpeace. The Amazon's silent crisis: Logging regulation & 5 ways to launder. Tech. Rep., Greenpeace Brazil (2014). Available at: <https://www.greenpeace.org/brazil/publications/report/2014/05/01/Amazon-s-silent-crisis-Logging-regulation-5-ways-to-launder>.

[//www.greenpeace.org/usa/wp-content/uploads/legacy/Global/usa/pla-net3/PDFs/Amazon5Ways.pdf](http://www.greenpeace.org/usa/wp-content/uploads/legacy/Global/usa/pla-net3/PDFs/Amazon5Ways.pdf).

- [18] Brancalion, P. H. *et al.* Fake legal logging in the brazilian amazon. *Science Advances* **4** (8), 1–8 (2018). <https://doi.org/10.1126/sciadv.aat1192>.
- [19] Valdiones, A. P. *et al.* A Evolução do setor madeireiro na Amazônia entre 1980 e 2020 e as oportunidades para o seu desenvolvimento inclusivo e sustentável na próxima década. Tech. Rep., SIMEX Network: Imazon, ICV, Imaflora, Idesam (2022). Available at: <https://imazon.org.br/publicacoes/a-evolucao-do-setor-madeireiro-na-amazonia-entre-1980-a-2020-e-as-oportunidades-para-seu-desenvolvimento-inclusivo-e-sustentavel-na-proxima-decada/>.
- [20] Dieter, M., Englert, H. & Weimar, H. Wood from illegal harvesting in EU Markets: Estimations and open issues. *Landbauforschung Volkenrode* **62** (4), 247–254 (2012) .
- [21] Hoare, A. Tackling illegal logging and the related trade: What progress and where next. Tech. Rep., Chatham House (2015). Available at: <https://www.chathamhouse.org/2015/07/tackling-illegal-logging-and-related-trade-what-progress-and-where-next>.
- [22] Macpherson, A. J., *et al.* The sustainability of timber production from Eastern Amazonian forests. *Land Use Policy* **29** (2), 339–350 (2012). <https://doi.org/10.1016/j.landusepol.2011.07.004>.
- [23] Hansen, C. P., Rutt, R. & Acheampong, E. ‘Experimental’ or business as usual? Implementing the European Union Forest Law Enforcement, Governance and Trade (FLEGT) Voluntary Partnership Agreement in

- Ghana. *Forest Policy and Economics* **96**, 75–82 (2018). <https://doi.org/10.1016/j.forpol.2018.08.012>.
- [24] Azevedo-Ramos, C., Silva, J. N. M. & Merry, F. The evolution of Brazilian forest concessions. *Elementa: Science of the Anthropocene* **3**, 000048 (2015). <https://doi.org/10.12952/journal.elementa.000048>.
- [25] Maryudi, A., *et al.* “A Level Playing Field”?—What an Environmental Justice Lens Can Tell us about Who Gets Leveled in the Forest Law Enforcement, Governance and Trade Action Plan. *Society and Natural Resources* **1920** (2020). <https://doi.org/10.1080/08941920.2020.1725201>.
- [26] Tritsch, I. *et al.* Multiple patterns of forest disturbance and logging shape forest landscapes in paragominas, brazil. *Forests* **7** (12) (2016). Available at: <https://www.mdpi.com/1999-4907/7/12/315>. <https://doi.org/10.3390/f7120315>.
- [27] Santos de Lima, L. *et al.* Illegal logging as a disincentive to the establishment of a sustainable forest sector in the amazon. *PLOS ONE* **13** (12), e0207855 (2018). <https://doi.org/10.1371/journal.pone.0207855>.
- [28] Celentano, D. *et al.* Desmatamento, degradação e violência no “Mosaico Gurupi” – A região mais ameaçada da Amazônia. *Estudos Avançados* **32** (92) (2018). <https://doi.org/10.5935/0103-4014.20180021>.
- [29] Jackson, B., *et al.* Understanding the co-occurrence of tree loss and modern slavery to improve efficacy of conservation actions and policies. *Conservation Science and Practice* **2** (5), e183 (2020). <https://doi.org/10.1111/csp2.183>.

- [30] IBAMA. Timber Production of Brazilian Native Species: 2012 to 2017 Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais Renováveis (IBAMA) (2019). Available at: <http://www.ibama.gov.br/flora-e-madeira/publicacoes>.
- [31] Schulze, M., *et al.* Evaluating ipê (Tabebuia, Bignoniaceae) logging in Amazonia: Sustainable management or catalyst for forest degradation? *Biological Conservation* **141** (8), 2071–2085 (2008). <https://doi.org/10.1016/j.biocon.2008.06.003>.
- [32] CITES. Notification to the parties: amendments of appendices I and II of the convention. Tech. Rep. 2022/052, Secretariat of the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES) (2022). Available at: <https://cites.org/sites/default/files/notifications/E-Notif-2022-052.pdf>.
- [33] Norman, M. & Zunino, A. R. Demand for luxury decks in europe and north america is pushing ipê to the brink of extinction across the amazon basin and threatening the forest frontier. Tech. Rep., Forest Trends (2022). Available at: <https://www.forest-trends.org/publications/demand-is-pushing-ipe-to-brink-of-extinction-across-the-amazon-basin/>.
- [34] Brazil. Projeto RADAM. Folha SB. 22 Araguaia e parte da folha SC. 22 Tocantins ; geologia, geomorfologia, solos, vegetação, uso potencial da terra. (1974). Available at: <https://biblioteca.ibge.gov.br/visualizacao/livros/liv24021.pdf>.
- [35] Andrade, F. W. C. *et al.* The Legal Roundwood Market in the Amazon and its impact on deforestation in the region between 2009–2015. *Forests* **13** (4), 558 (2022). <https://doi.org/10.3390/f13040558>, number: 4



Publisher: Multidisciplinary Digital Publishing Institute.

- [36] TCU. Relatório de monitoramento TC Nº 004.579/2011-9. Tech. Rep., Tribunal de Contas da União (TCU) - 8ª Secretaria de Controle Externo (2011). Available at: <https://portal.tcu.gov.br/biblioteca-digital/auditoria-operacional-no-sistema-de-controle-de-transito-de-produtos-florestais.htm>.
- [37] TCU. Relatório de monitoramento TC Nº 018.352/2015-4. Tech. Rep., Tribunal de Contas da União (TCU) - Secretaria-Geral de Controle Externo Secretaria de Controle Externo da Agricultura e do Meio Ambiente (2015). Available at: <https://portal.tcu.gov.br/biblioteca-digital/auditoria-operacional-no-sistema-de-controle-de-transito-de-produtos-florestais.htm>.
- [38] Conceição, K. V. *et al.* Government policies endanger the indigenous peoples of the Brazilian Amazon. *Land Use Policy* **108**, 105663 (2021). <https://doi.org/10.1016/j.landusepol.2021.105663>.
- [39] Börner, J. *et al.* Post-crackdown effectiveness of field-based forest law enforcement in the brazilian amazon. *PLOS ONE* **10** (4), e0121544 (2015). <https://doi.org/10.1371/journal.pone.0121544>.
- [40] Hummel, A. C. Deforestation in the amazon: What is illegal and what is not? *Elementa: Science of the Anthropocene* **4**, 000141 (2016). <https://doi.org/10.12952/journal.elementa.000141>.
- [41] Milhorange, C. Policy dismantling and democratic regression in brazil under bolsonaro: Coalition politics, ideas, and underlying discourses. *Review of Policy Research* **39** (6), 752–770 (2022). <https://doi.org/10.1>

[111/ropr.12502](#).

- [42] Sotirov, M., *et al.* Policy options to regulate timber and agricultural supply-chains for legality and sustainability: The case of the EU and Brazil. *Forest Policy and Economics* **144**, 102818 (2022). <https://doi.org/10.1016/j.forpol.2022.102818>.
- [43] Brazil. IBAMA Normative Instruction N<sup>o</sup> 3 23.01.2020 (2020). Available at: <https://www.in.gov.br/web/dou/-/instrucao-normativa-n-3-de-23-de-janeiro-de-2020-240087843>. ISBN: 2020240087843.
- [44] Brazil. IBAMA Normative Instruction N<sup>o</sup> 16 25.11.2022 (2022). Available at: <https://www.in.gov.br/web/dou/-/instrucao-normativa-n-16-de-25-de-novembro-de-2022-448030474>.
- [45] Brazil. Law N<sup>o</sup> 12.651 25.05.2012 (2012). Available at: [http://www.planalto.gov.br/ccivil\\_03/\\_Ato2011-2014/2012/Lei/L12651.htm](http://www.planalto.gov.br/ccivil_03/_Ato2011-2014/2012/Lei/L12651.htm).
- [46] Brazil. Law N<sup>o</sup> 9.605, 12.02.1998 (1998). Available at: [http://www.planalto.gov.br/ccivil\\_03/leis/L9605.htm](http://www.planalto.gov.br/ccivil_03/leis/L9605.htm).
- [47] Brazil. Portaria (Normative Rule) IBDF (Instituto Brasileiro de Desenvolvimento Florestal) N<sup>o</sup> 486 28.11.1986 (1986). Available at: [www.ibama.gov.br/sophia/cnia/legislacao/IBDF/PT0486-281086.PDF](http://www.ibama.gov.br/sophia/cnia/legislacao/IBDF/PT0486-281086.PDF).
- [48] Brazil. CONAMA Resolution N<sup>o</sup> 406 02.02.2009 (2009). Available at: <http://www.ibama.gov.br/sophia/cnia/legislacao/MMA/RE0406-020209.PDF>.
- [49] Brazil. MMA (Ministério do Meio Ambiente) Normative Instruction N<sup>o</sup> 1 12.02.2015 (2015). Available at: [www.ibama.gov.br/sophia/cnia/legis](http://www.ibama.gov.br/sophia/cnia/legis)

[lacao/MMA/IN0001-12022015.pdf](#).

- [50] Richardson, V. A. & Peres, C. A. Temporal Decay in Timber Species Composition and Value in Amazonian Logging Concessions. *PLOS ONE* **11** (7), e0159035 (2016). <https://doi.org/10.1371/journal.pone.0159035>.
- [51] Greenpeace. Imaginary trees, real destruction. Tech. Rep., Greenpeace Brazil (2018). Available at: [https://www.greenpeace.org/static/planet4-international-stateless/2018/03/b91d03c3-greenpeace-report\\_imaginary-trees-real-destruction\\_march-2018.pdf](https://www.greenpeace.org/static/planet4-international-stateless/2018/03/b91d03c3-greenpeace-report_imaginary-trees-real-destruction_march-2018.pdf).
- [52] Brancalion, P. *et al.* Fake legal logging in the brazilian amazon (2018). Zenodo [10.5281/zenodo.1244107](https://doi.org/10.5281/zenodo.1244107).
- [53] Kastner, T., Kastner, M. & Nonhebel, S. Tracing distant environmental impacts of agricultural products from a consumer perspective. *Ecological Economics* **70** (6), 1032–1040 (2011). <https://doi.org/10.1016/j.ecolecon.2011.01.012>.
- [54] Brazil. CONAMA Resolution N<sup>o</sup> 474 06.04.2016 (2016). Available at: <http://www2.mma.gov.br/port/conama/legiabre.cfm?codlegi=720>.
- [55] da Silva Luz, E. *et al.* Challenges of the lumber production in the Amazon region: relation between sustainability of sawmills, process yield and logs quality. *Environment, Development and Sustainability* **23** (4), 4924–4948 (2021). <https://doi.org/10.1007/s10668-020-00797-9>, publisher: Springer Netherlands.
- [56] West, T. A. & Fearnside, P. M. Brazil’s conservation reform and the reduction of deforestation in Amazonia. *Land Use Policy* **100** (August 2020) (2021). <https://doi.org/10.1016/j.landusepol.2020.105072>.

- [57] Romero, F. M. B. *et al.* Stocks of carbon in logs and timber products from forest management in the southwestern amazon. *Forests* **11** (10), 1–20 (2020). <https://doi.org/10.3390/f11101113>.
- [58] Brazil. CONAMA Resolution N<sup>o</sup> 497 19.08.2020 (2020). Available at: <https://www.in.gov.br/en/web/dou/-/resolucao-n-497-de-19-de-agosto-de-2020-273217612>.
- [59] Brazil. Law N<sup>o</sup> 10.650, 16.04.2003. (2003). Available at: [https://www.planalto.gov.br/ccivil\\_03/leis/2003/110.650.htm](https://www.planalto.gov.br/ccivil_03/leis/2003/110.650.htm).
- [60] SFB. Plano Anual de Outorga Florestal - PAOF 2022 (2021). Available at: [https://www.gov.br/agricultura/pt-br/assuntos/servico-florestal-brasileiro/concessao-florestal/plano-anual-de-outorga-florestal/Paof\\_2022\\_\\_\\_versao\\_27\\_de\\_julho\\_20211.pdf](https://www.gov.br/agricultura/pt-br/assuntos/servico-florestal-brasileiro/concessao-florestal/plano-anual-de-outorga-florestal/Paof_2022___versao_27_de_julho_20211.pdf).
- [61] Grose, S. O. & Olmstead, R. G. Taxonomic Revisions in the Polyphyletic Genus *Tabebuia* s. l. (Bignoniaceae). *Systematic Botany* **32** (3), 660–670 (2007). Available at: <http://www.jstor.org/stable/25064275>.
- [62] Tejada, G., Görgens, E. B., Espírito-Santo, F. D. B., Cantinho, R. Z. & Ometto, J. P. Evaluating spatial coverage of data on the aboveground biomass in undisturbed forests in the Brazilian Amazon. *Carbon Balance and Management* **14** (1), 11 (2019). <https://doi.org/10.1186/s13021-019-0126-8>.
- [63] GovPA. Normative Instruction SEMA N<sup>o</sup> 1 10.03.2008 (2008). Available at: <https://www.sistemas.pa.gov.br/sisleis/legislacao/2564>.

- [64] Elmagarmid, A. K., Ipeirotis, P. G. & Verykios, V. S. Duplicate Record Detection: A Survey. *IEEE Transactions on Knowledge and Data Engineering* **19** (1), 1–16 (2007). <https://doi.org/10.1109/TKDE.2007.250581>.
- [65] Brazil. Law Nº 13.709, 14.08.2018 (2018). Available at: [https://www.planalto.gov.br/ccivil\\_03/\\_ato2015-2018/2018/lei/l13709.htm](https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm).
- [66] R Core Team. R: A Language and Environment for Statistical Computing R Foundation for Statistical Computing (2023). Available at: <https://www.R-project.org/>.
- [67] Pereira, R. H. M. & Goncalves, C. N. *geobr: Download Official Spatial Data Sets of Brazil* (2022). Available at: <https://CRAN.R-project.org/package=geobr>. R package version 1.7.0.
- [68] Wickham, H. *et al.* Welcome to the tidyverse. *Journal of Open Source Software* **4** (43), 1686 (2019). <https://doi.org/10.21105/joss.01686> .
- [69] Firke, S. *janitor: Simple Tools for Examining and Cleaning Dirty Data* (2023). Available at: <https://CRAN.R-project.org/package=janitor>. R package version 2.2.0.
- [70] Pebesma, E. & Bivand, R. *Spatial Data Science: With applications in R* (Chapman and Hall/CRC, 2023). Available at: <https://r-spatial.org/book/>.
- [71] Tennekes, M. tmap: Thematic maps in R. *Journal of Statistical Software* **84** (6), 1–39 (2018). <https://doi.org/10.18637/jss.v084.i06>.

- [72] Maechler, M. & Dutang, C. & Goulet, V. *expm: Matrix Exponential, Log, 'etc'.* (2023). Available at: <https://CRAN.R-project.org/package=expm>. R package version 0.999-7.
- [73] Aphalo, P. J. *ggpmisc: Miscellaneous Extensions to 'ggplot2'* (2022). Available at: <https://CRAN.R-project.org/package=ggpmisc>. R package version 0.5.2.
- [74] Wickham, H. Reshaping data with the reshape package. *Journal of Statistical Software* **21** (12), 1–20 (2007). Available at: <http://www.jstatsoft.org/v21/i12/>.
- [75] Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer-Verlag New York, 2016). Available at: <https://ggplot2.tidyverse.org>.
- [76] Yu, G. *scatterpie: Scatter Pie Plot* (2023). Available at: <https://CRAN.R-project.org/package=scatterpie>. R package version 0.1.9.
- [77] Brunson, J. C. ggalluvial: Layered grammar for alluvial plots. *Journal of Open Source Software* **5** (49), 2017 (2020). <https://doi.org/10.21105/joss.02017>.
- [78] Pedersen, T. L. *patchwork: The Composer of Plots* (2022). Available at: <https://CRAN.R-project.org/package=patchwork>. R package version 1.1.2.
- [79] Wickham, H. & Seidel, D. *scales: Scale Functions for Visualization* (2022). Available at: <https://CRAN.R-project.org/package=scales>. R package version 1.2.1.

- [80] Hvitfeldt, E. *paletteer: Comprehensive Collection of Color Palettes* (2021). Available at: <https://github.com/EmilHvitfeldt/paletteer>. R package version 1.3.0.
- [81] Pedersen, T. L. & Crameri, F. *scico: Colour Palettes Based on the Scientific Colour-Maps* (2022). Available at: <https://CRAN.R-project.org/package=scico>. R package version 1.3.1.