

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- ☐ ☒ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- ☐ ☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- ☒ ☐ The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- ☒ ☐ A description of all covariates tested
- ☐ ☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- ☐ ☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- ☒ ☐ For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- ☒ ☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- ☒ ☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- ☒ ☐ Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection	The automated download of SINAFLOR data was carried out in R (last replicated in R 4.3.1). Administrative boundaries were obtained via API from geobr R library. The automated download of logging permits and the scraping of logging permit PDFs were carried out in Python (last replicated in Python 3.10 and 3.11 respectively). As of March 2023, the SIMLAM-SEMAS webpage used in the auto-download of logging permit PDFs underwent a major update and logging permits data can now be accessed via http://portaldatransparencia.semas.pa.gov.br/ .
Data analysis	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) and made use of several open source libraries. Data cleaning, wrangling and analyses used particularly the tidyverse, janitor, sf, base R, but also tmap, expm, ggpmisc, reshape2. Graphs displayed throughout the manuscripts were based particularly on ggplot2, scatterpie and ggalluvial, but also on patchwork, scales, paletteer and scico libraries. The code necessary to replicate the analyses presented in this study are publicly available at https://github.com/carolsrto/illegality-risk-ns .

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

The data necessary to replicate the analyses presented in this study are publicly available on Zenodo <https://doi.org/10.5281/zenodo.8068431>. All data used in this study is in the public domain in line with Law 10.650/2003 on SISNAMA public data access, nonetheless, not in all instances readily accessible given the transparency caveats discussed in this paper. The logging permits data under state jurisdiction were originally accessed through Pará's Environmental Secretariat SIMLAM at <http://monitoramento.semas.pa.gov.br/simlam/index.htm>. The webpage underwent a major update on March, 2023 with the launch of a new Transparency Portal and the link redirect users to <http://portal.datatransparencia.semas.pa.gov.br/>, where this state-level logging permit data can now be found. The logging permits data under national jurisdiction were accessed through the Brazilian Open Data Portal at <https://dados.gov.br/dados/conjuntos-dados/dof-autorizacoes-de-exploracao-florestal>. The timber transport data from SISFLORA-PA GFs were accessed via the Timberflow initiative led by Imaflora <http://timberflow.org.br/>. The timber transport data from SINAFLO DOFs were accessed through the Brazilian Open Data Portal at <https://dados.gov.br/dados/conjuntos-dados/dof-transportes-de-produtos-florestais>. The RADAM national inventory data was accessed through Brancalion et al. 2018 Zenodo repository. All other supporting data sources have been referenced.

Human research participants

Policy information about [studies involving human research participants and Sex and Gender in Research](#).

Reporting on sex and gender

N/A

Population characteristics

N/A

Recruitment

N/A

Ethics oversight

N/A

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description

We quantify sources of illegality risk and indicate at what stage in the supply-chain illegally logged timber is potentially laundered. First, we quantify the share of timber with missing or invalid logging permits being commercialized by matching timber transactions data to logging permits. Second, building on Brancalion et al. 2018, we quantify the practice of overstating yields, comparing stated harvest volumes per area authorized for logging with stocking rates from the RADAM national forest inventory. Lastly, we compare inflows and outflows of timber by converting all relevant ipê products to volumes of roundwood-equivalent. The three quantified sources of illegality risk are traced downstream to apparent consumption through an actor-level, physical input-output model that overcomes the lack of full traceability, allowing us to quantify the upstream embodied illegality risks associated with the consumption of ipê originating from Pará.

Research sample

We used all data available for logging permits and timber transactions relevant for the scope of the analysis. For the RADAM forest inventory, we used all data from the 1-ha RADAM plots (DBH, ≥ 50 cm) in line with the replication of Brancalion et al. 2018 study. We used a total of 394 unique RADAM plots.

Sampling strategy

No sampling strategy was applied. We reproduced the results of Brancalion et al. 2018 for ipê yields estimate.

Data collection

We collected and compiled data from different sources and databases, all referenced throughout the study.

Timing and spatial scale

The study covers the period between 2009-2019 and focuses on the Brazilian state of Pará.

Data exclusions

No data were excluded from the analysis. Any data points removed from visualizations are made reference to in captions and text.

Reproducibility

Custom code was developed to allow for the full reproducibility of the current study, including data collection (when applicable), cleaning, integration and analyses as well as output visualization and mapping. To be sensitive in what pertains to data protection laws, not all intermediary code (e.g. cleaning and integration steps) are made available in the public repository, but can be requested directly to authors. The study can also be fully reproduced by obtaining the same datasets and following the methodological steps described in methods and supplementary information.

Randomization

In determining baseline figures for naturally occurring ipê yields, and because the RADAM plots are 1ha and authorized logging areas were significantly larger, we performed 10,000 randomisation's of 19 1-ha plots (minimum authorized area) in order to build a comparable density distribution. This species yields estimation was replicated from Brancalion et al 2018 study approach.

Blinding

Not applicable to our study.

Did the study involve field work?

☐ Yes

☒ No

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging