Bibliometric analysis of TreesLab scientific production

Alber Sánchez alber.ipia@inpe.br
Guilherme Mataveli guilherme.mataveli@inpe.br

Research assistant - TreesLab
National Institute for Space Research - INPE
Brazil

November 23, 2023

Overview

Introduction

- Most of the definitions in this presentation were taken from Aria and Cuccurullo or any of their papers, webpages, tutorials, or videos.
- We assumed that TreesLab's papers are an study subject.

Bibliometrix package

- R package for bibliometric analysis [1].
- It allows quantitative research in bibliometrics and scientometrics.
- Statistical analysis of publications.
- Useful for performance evaluation and policy making.

Bibliometrix

- Import and convert data from bibliographic databases.
- Analysis of a publication dataset.
- Building matrices of co-citation, coupling, collaboration, and co-word analysis.

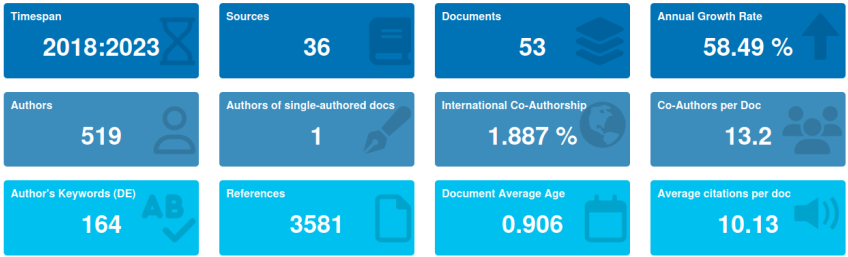
Bibliographic databases

- Scopus.
- Web of science.
- To query them, use a machine at INPE and your Café login.

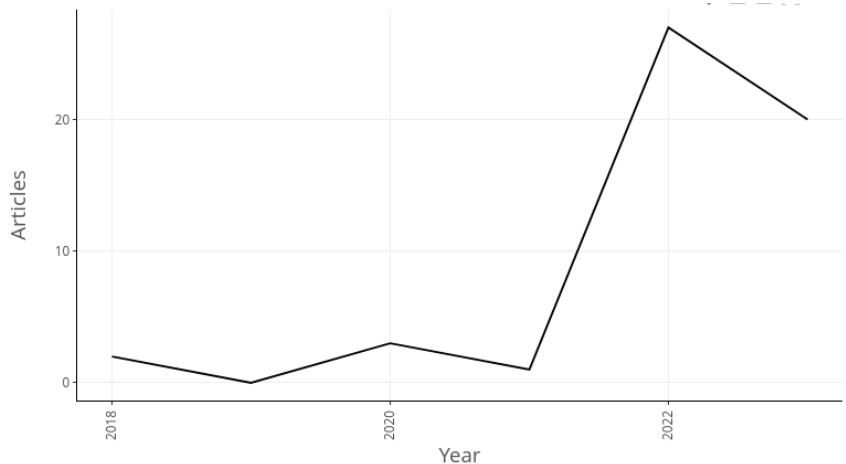
Data pre-processing

1. Retrieve DOI from TreesLab publication's webpage.
2. Query Scopus .
3. Query Web of Science.
4. Merge query results (Bibliometrix).
5. Run analysis .
   - Bibliometrix: R coders.
   - Biblioshiny: Non-coders.

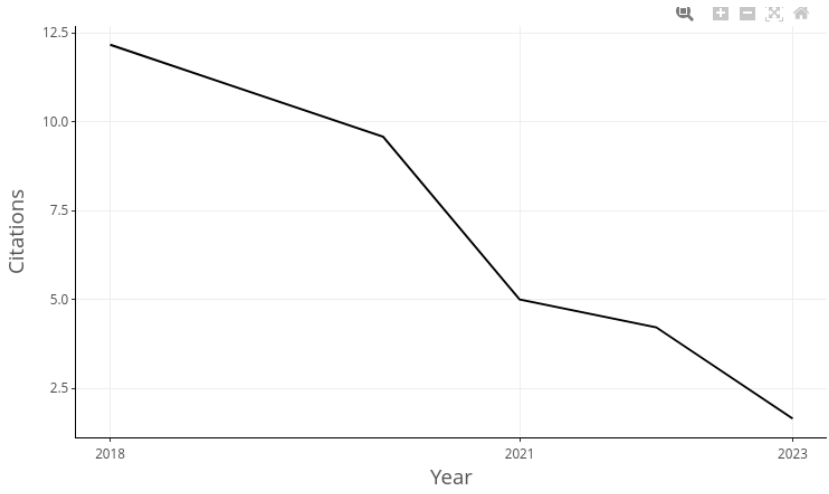
Main information



Annual scientific production



Average citations per year

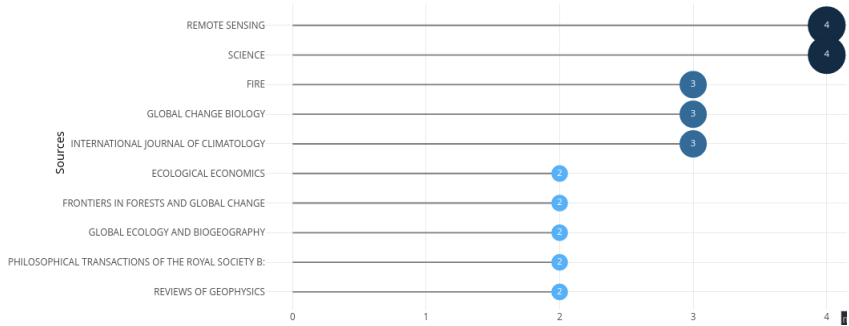


Three fields plot

This is an interactive figure (See live). Possible combinations:

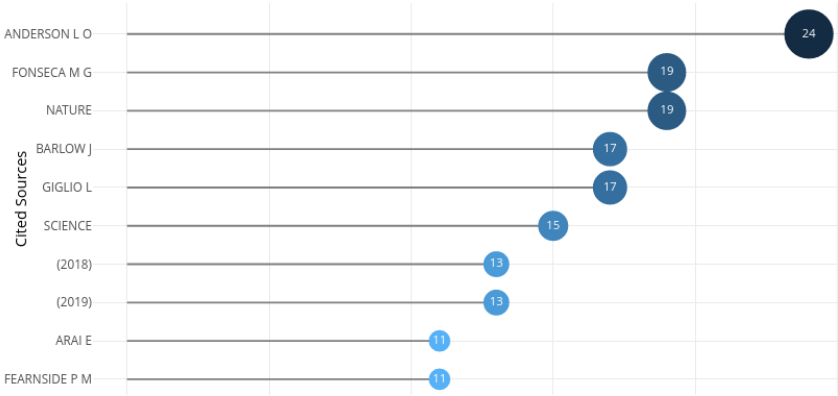
- Top keywords, authors, and journals.
- Top authors, references they cite, and keywords they use.

Most relevant sources



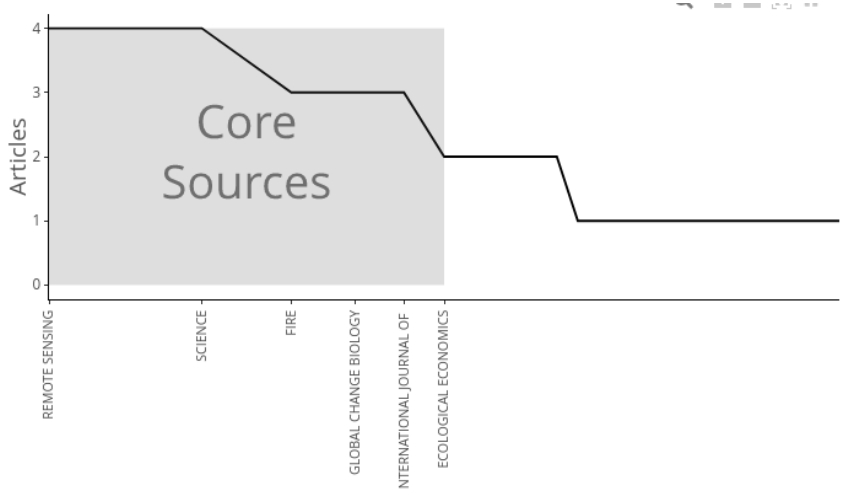
A *source* is a journal/book/proceeding/series/etc. which published one or more documents included in the bibliographic collection.

Most local cited sources



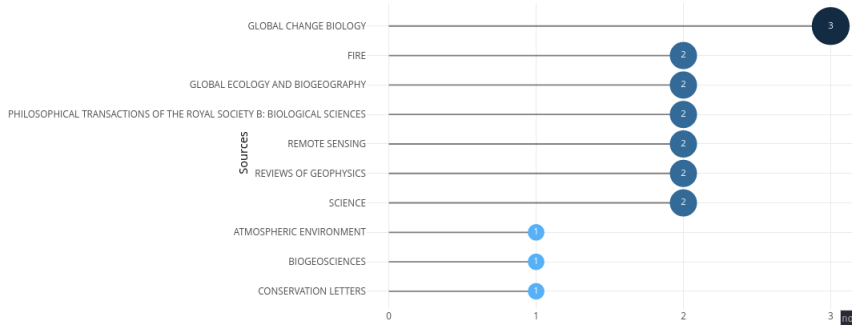
A *cited source* is a journal/book/proceeding/series/etc. included in at least one of the reference list (bibliography) of the document set.

Bradford's law



The journals particularly devoted to TreesLab's subjects.

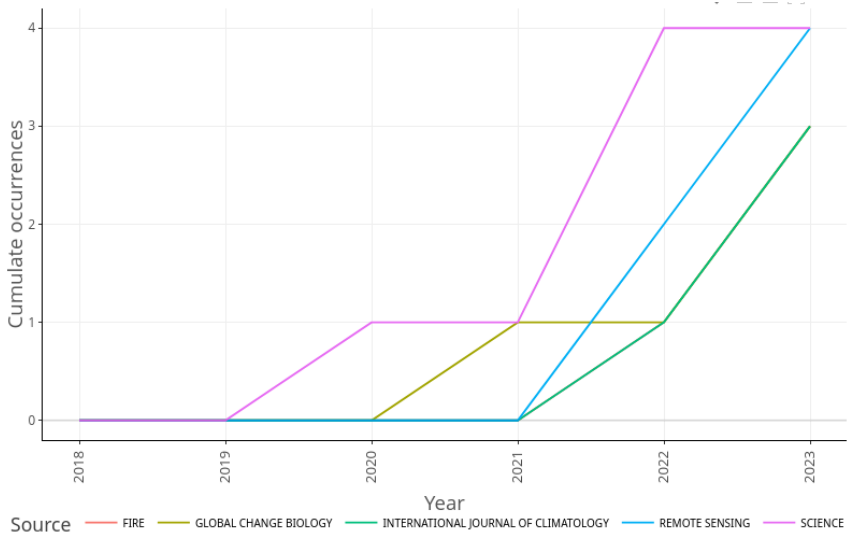
Source's local impact



Impact by H-Index and its generalizations.

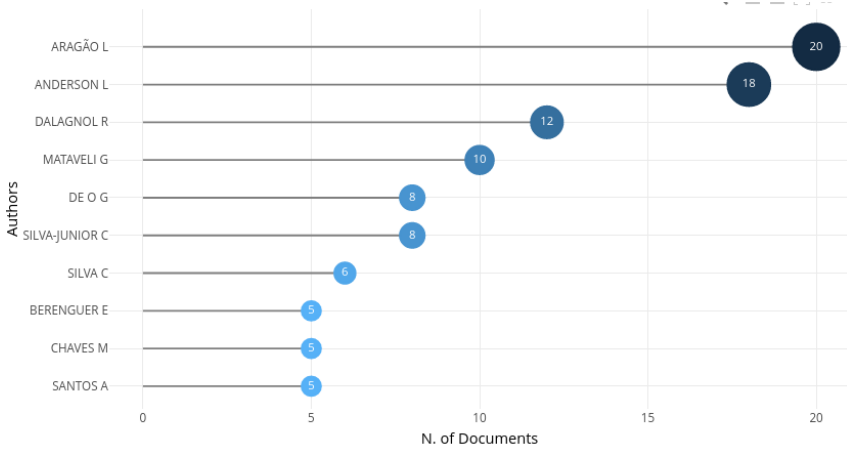
The Hirsh index (*H-index*) is an author's (or journal's) number of published articles (h) each of which has been cited in other papers at least h times.

Sources' production over time



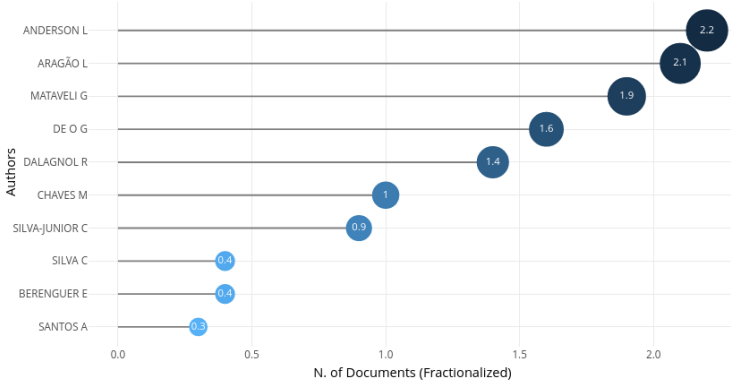
Number of publications per year.

Most relevant authors



Most relevant authors (number of documents).

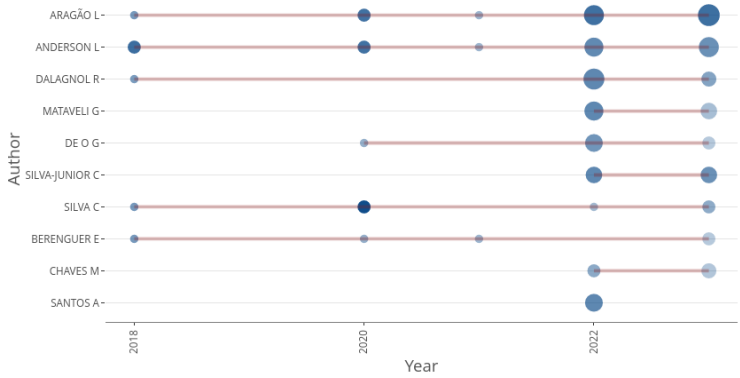
Most relevant authors



Most relevant authors (fractionalized).

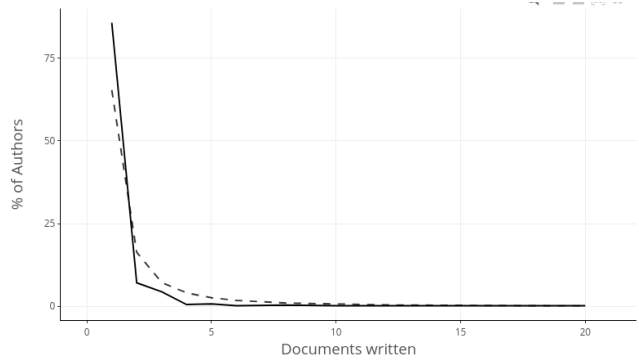
Fractionalized authorship quantifies an individual author's contributions to a published set of papers (following the hypothesis of uniform contributions of all co-authors at each document).

Author's production over time



A line represents an author's timeline. The buble size is proportional to the number of documents. The color intensity is proportional to the total citations per year.

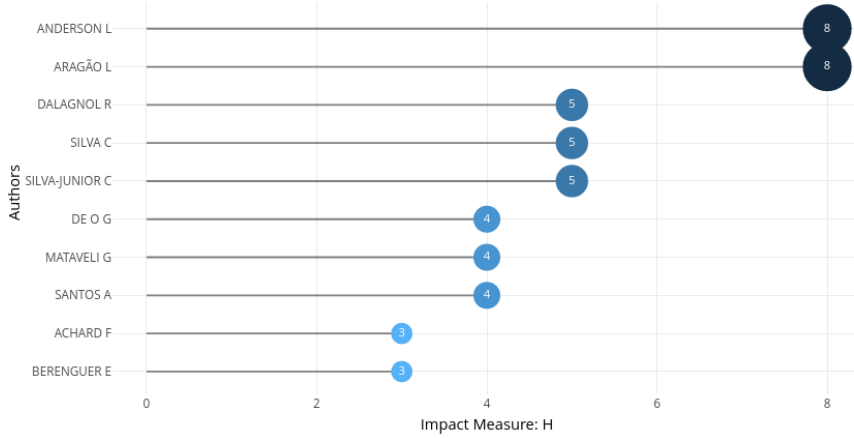
Author productivity through Lotka's law



Dashed line represents the theoretical distribution.

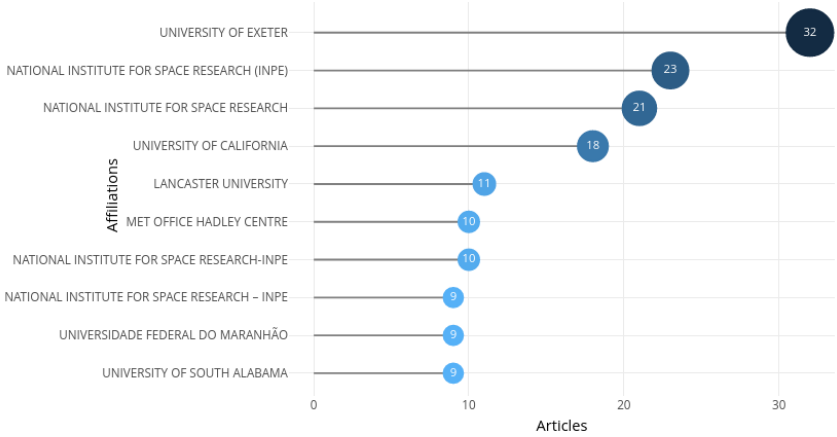
As the number of articles published increases, authors producing that many publications become less frequent.

Authors' local impact



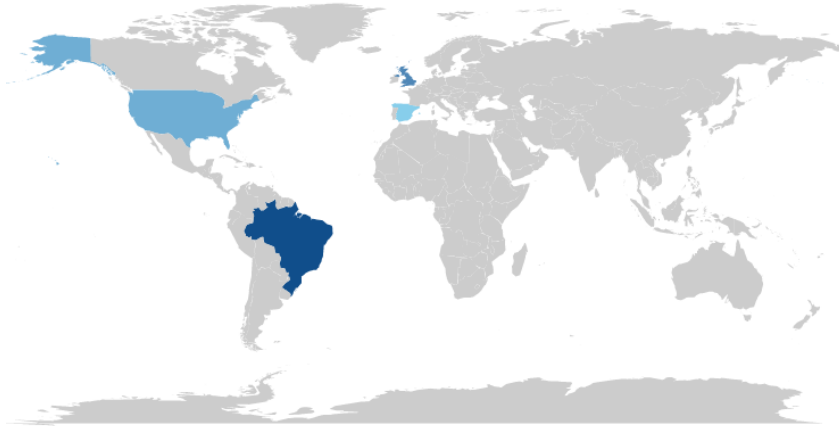
Authors' local impact by H-index and its generalizations.

Most relevant affiliations



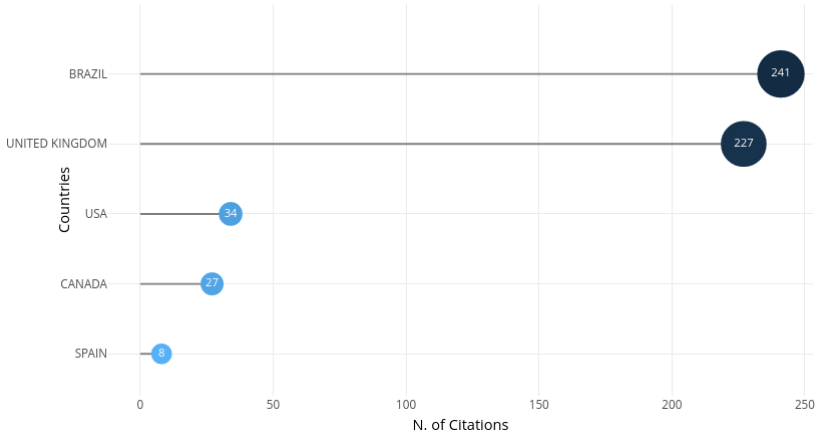
INPE's affiliation strings need cleaning!

Countries' scientific production



The color intensity is proportional to the number of publications.

Most cited countries



Documents and references

- *Document (or citing document)*: Scientific document (article, review, conference proceeding, etc.) included in a bibliographic collection.
- *Reference (or cited reference)*: Scientific document included in at least one of the reference lists (bibliography) of the document set. Then "a reference is cited by one or more documents".
- *Cited document*: Scientific document included in a bibliographic collection and, at the same time, it is cited in at least one other document in the collection. Cited documents are a subset of the reference set.

Global and local citations

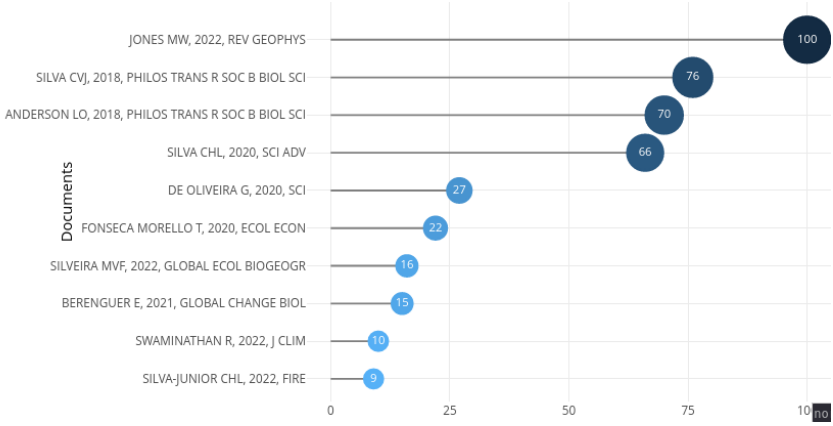
Global citations.

- Measures the number of citations a document has received from documents contained in the entire database (e.g. WoS or Scopus).
- Measures the impact of a document in the whole bibliographic database.
- For many documents, a large part of global citations could come from other disciplines!

Local citations.

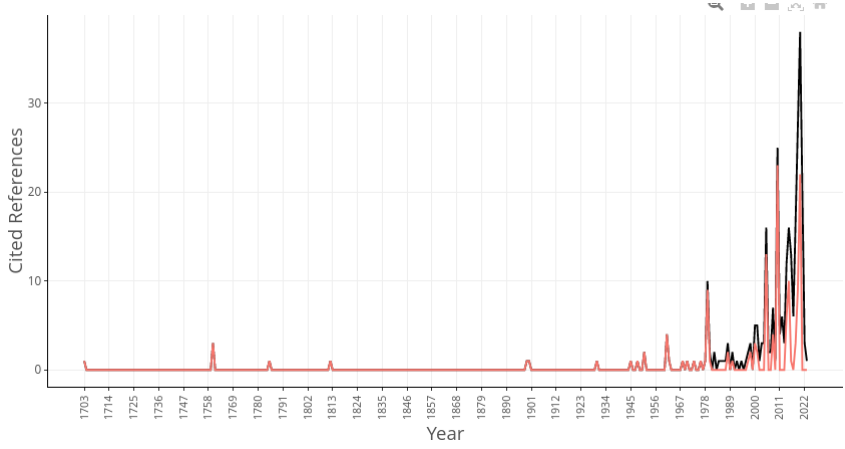
- Measures the number of citations a document has received from documents included in the analyzed collection.
- Is calculated analyzing the whole reference set.
- Measures the impact of a document in the analyzed collection.

Most global cited documents



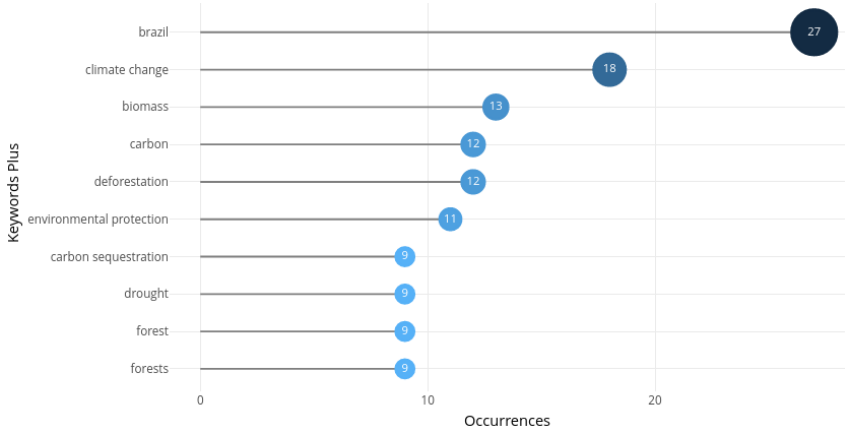
Is there an impactful paper?

Reference publication year spectroscopy



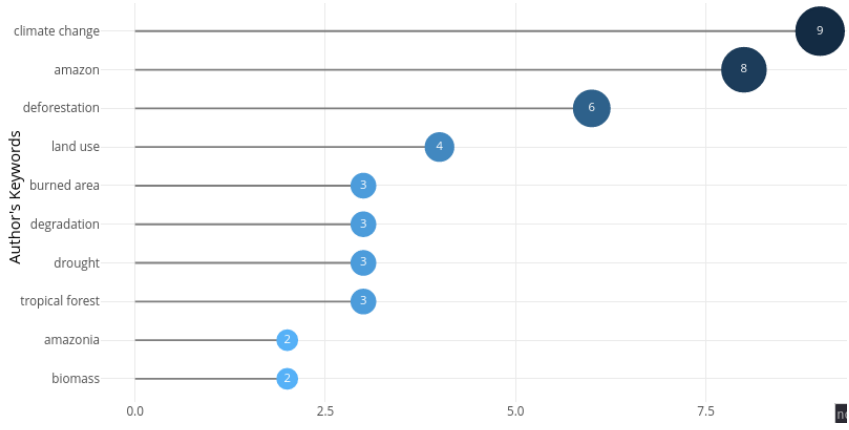
Number of cited references per year and deviation from the 5-year mean.

Most frequent words (Keywords Plus)



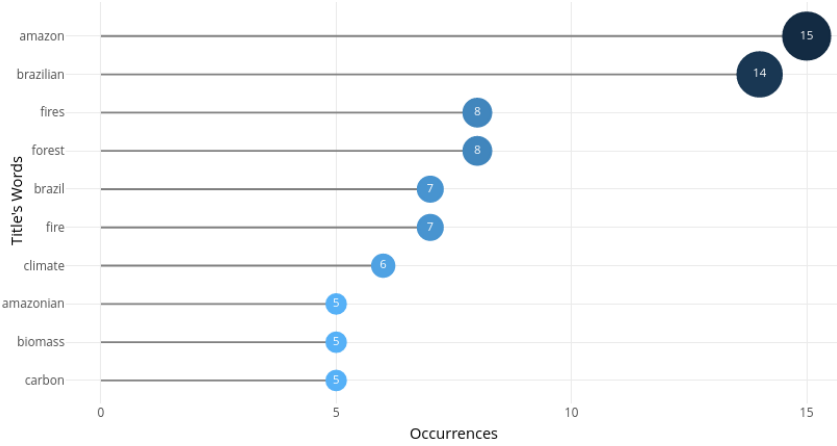
Keywords Plus are words or phrases that appear frequently in the titles of an article's references and not necessarily in the title of the article or as Author Keywords.

Most frequent words (Author's keywords)



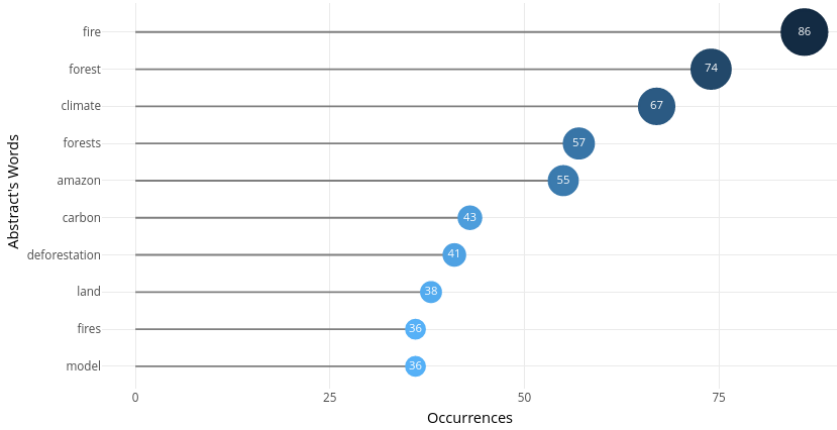
Authors' Keywords are the terms the authors believe best represent the content of their papers. Beware of plurals and conjugations.

Most frequent words (titles' words)



Words extracted from titles (or abstracts) removing "stop words" and punctuation.

Most frequent words (abstracts' words)



Abstract words need to be cleaned to avoid trivial terms such as "paper", "study", "work", "data", etc.

Wordclouds

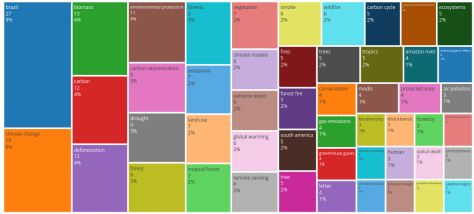
(a) Keywords Plus

(b) Authors' Keywords

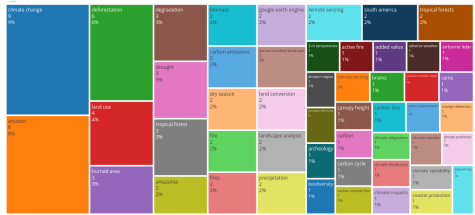
(c) Titles' words.

(d) Abstracts' words.

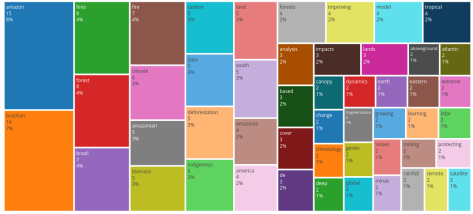
Treemaps



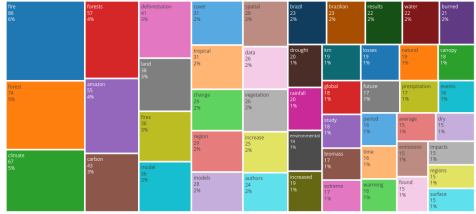
(a) Keywords Plus



(b) Authors' Keywords

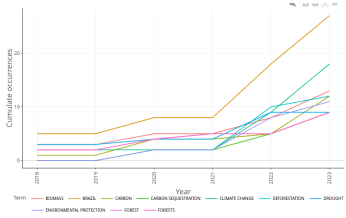


(c) Titles' words.

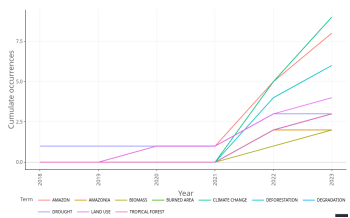


(d) Abstracts' words.

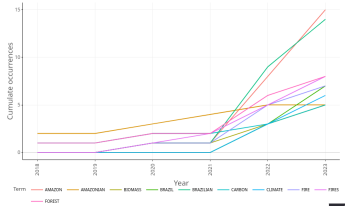
Words' frequency over time



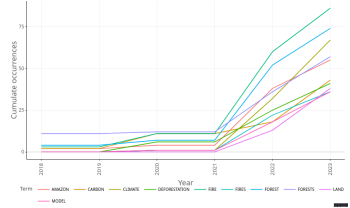
(a) Keywords Plus



(b) Authors' Keywords



(c) Titles' words.



(d) Abstracts' words.

Structures of knowledge

- *Conceptual*: What science talks about.
- *Intellectual*: How the work of an author influences a given scientific community.
- *Social*: How authors, institutions, and countries interact each other.

Science mapping allows investigating scientific knowledge from a statistical point of view.

Take home message

- TODO.

References I

[1] Massimo Aria and Corrado Cuccurullo. Bibliometrix : An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, 11(4):959–975, November 2017. ISSN 17511577. doi: 10.1016/j.joi.2017.08.007.

Queries

Queries for TreesLab's papers.

Query Scopus I

```
DOI("10.1111/gcb.16670") OR
("10.1016/j.atmosenv.2022.118954") OR
("10.3390/rs14164092") OR
("10.1111/conl.12908.") OR
("10.1126/science.ade6396") OR
("10.1111/geb.13577") OR
("10.1080/2150704X.2022.2109942") OR
("10.3389/fenvs.2022.946729") OR
("10.1080/01431161.2022.2106801") OR
("10.3390/rs14071545") OR
("10.1002/rse2.264") OR
("10.3390/fire5030077") OR
("10.1016/j.ecolecon.2019.106501") OR
("10.3390/fire6110426") OR
("10.1126/science.ade2541") OR
```

Query Scopus II

```
16  ("10.1016/j.ecolecon.2023.107983") OR
17  ("10.3389/fphy.2022.1064162") OR
18  ("10.3389/ffgc.2023.1107417") OR
19  ("10.1038/s41598-023-32746-7") OR
20  ("10.3389/ffgc.2023.1024101") OR
21  ("10.1038/s41586-022-05679-w") OR
22  ("10.1111/gcb.16670") OR
23  ("10.3390/fire6070275") OR
24  ("10.3390/fire6010002") OR
25  ("10.1111/gcb.16513") OR
26  ("10.1111/1365-2745.14003") OR
27  ("10.3390/rs14071545") OR
28  ("10.1016/j.ecolmodel.2021.109817") OR
29  ("10.1080/23754931.2022.2117565") OR
30  ("10.1016/j.atmosenv.2022.118954") OR
```

Query Scopus III

```
("10.3390/f13010016") OR
("10.3390/rs14164092") OR
("10.1111/conl.12908") OR
("10.1126/science.abn4936") OR
("10.1126/science.ade6396") OR
("10.1111/geb.13577") OR
("10.1080/2150704X.2022.2109942") OR
("10.1080/01431161.2022.2106801") OR
("10.1002/rse2.264") OR
("10.1111/geb.13563") OR
("10.1126/science.abo4578") OR
("10.1002/cli2.19") OR
("10.1002/cli2.15") OR
("10.1002/cli2.11") OR
("10.1002/cli2.8") OR
```

Query Scopus IV

```
46  ("10.1111/gcb.15425") OR
47  ("10.1016/j.ecolecon.2019.106501") OR
48  ("10.1126/sciadv.aaz8360") OR
49  ("10.1126/science.abd5942") OR
50  ("10.1098/rstb.2017.0411") OR
51  ("10.1098/rstb.2018.0043")
```

Query WoS I

```
DO=(10.1111/gcb.16670
10.1016/j.atmosenv.2022.118954
10.3390/rs14164092
10.1111/conl.12908.
10.1126/science.ade6396
10.1111/geb.13577
10.1080/2150704X.2022.2109942
10.3389/fenvs.2022.946729
10.1080/01431161.2022.2106801
10.3390/rs14071545
10.1002/rse2.264
10.3390/fire5030077
10.1016/j.ecolecon.2019.106501
10.3390/fire6110426
10.1126/science.ade2541
```

Query WoS II

```
10.1016/j.ecolecon.2023.107983
10.3389/fphy.2022.1064162
10.3389/ffgc.2023.1107417
10.1038/s41598-023-32746-7
10.3389/ffgc.2023.1024101
10.1038/s41586-022-05679-w
10.1111/gcb.16670
10.3390/fire6070275
10.3390/fire6010002
10.1111/gcb.16513
10.1111/1365-2745.14003
10.3390/rs14071545
10.1016/j.ecolmodel.2021.109817
10.1080/23754931.2022.2117565
10.1016/j.atmosenv.2022.118954
```

Query WoS III

```
10.3390/f13010016
10.3390/rs14164092
10.1111/conl.12908
10.1126/science.abn4936
10.1126/science.ade6396
10.1111/geb.13577
10.1080/2150704X.2022.2109942
10.1080/01431161.2022.2106801
10.1002/rse2.264
10.1111/geb.13563
10.1126/science.abo4578
10.1002/cli2.19
10.1002/cli2.15
10.1002/cli2.11
10.1002/cli2.8
```

Query WoS IV

```
46  10.1111/gcb.15425
47  10.1016/j.ecolecon.2019.106501
48  10.1126/sciadv.aaz8360
49  10.1126/science.abd5942
50  10.1098/rstb.2017.0411
51  10.1098/rstb.2018.0043)
```