A Topography of Climate Change Research

Max Callaghan[1,2]

[1]Mercator Research Institute on Global Commons and Climate Change, Torgauer Straße, 10829 Berlin, Germany
[2]School of Earth and Environment, University of Leeds, Leeds LS2 9JT, United Kingdom

Draft current July 16, 2018

The massive expansion of scientific literature on climate change challenges the Intergovenmental Panel on Climate Change (IPCC)'s ability to assess the science according to its objectives. Moreover, the number and variety of papers hinders researchers of the science-policy interface from making objective judgements about those IPCC assessments. In this paper, we present a novel application of a machine-reading approach to model the topical content of papers on climate change. This dynamic topic model provides the basis for a *topography* of climate change literature. The thematic development of the field is outlined and used to inform an analysis of the topics which are better and less well covered by IPCC reports.

To deal with the wicked problem of climate change, international policy-makers need the IPCC. The IPCC as map-makers.

The IPCC sees its role as to "assess on a comprehensive, objective, open and transparent basis the scientific, technical and socio-economic information relevant to [...] climate change" [1]. Climate science is so broad, multi-disciplinary, and laden with uncertainties and values, that the role of the IPCC as assessment maker is vitally important to developing evidence-based international climate policy. Making maps [2]

The task of the IPCC has become much more difficult with big literature

Further, it has been pointed out that, in the age of "big literature", providing assessments that are comprehensive, objective and transparent has become much more difficult [3].

When IPCC's citations constitute an ever-decreasing proportion of the totality of science on climate change, questions about the map that the IPCC reports produce become more pressing:

- Is the map up to date? Is it complete? Is the perspective representative?

The IPCC, its reports and processes have been the object of study before. These are also hampered by problems of scale though

Various researchers have attempted to do empirical research on the assessment reports, and processes of inter. alia. the IPCC [4] [5] [6].

Policy makers, when asked about their interactions with the IPCC call for a greater focus on solutions [7]

These studies are similarly challenged by the the size of the literature. Traditional bibliometric techniques are insufficient.

Some literature exists on bibliometrics and climate change, but tends not to deal with text

Bibliometrics e.g. [8] [9]

Text based approaches are usually of a smaller scope [10] or methodological contributions [11]

The scale of the problem in context

The scale of the challenge is depicted in figure 1. Less than two thousand documents relevant to climate change were published before the first assessment report (see Methods for data, exclusions and processing). These documents contained 3,528 unique terms, each of which was used on average in 0.49% of documents. In the three complete years since the publication of AR5, 128,357 documents have been published, containing 86,419 unique terms, used on average in 0.12% of documents. To put this into context, the 1,189 chapters of the Bible contain a vocabulary of 11,977 unique words. Put another way, the 236,634 publications published in AR5 and AR6 are significantly larger than the 178,118 publications recorded in the first volume of the 'Catalogue of Scientific Papers', compiled by the Royal Society to record the entirety of scientific output from 1800 to 1863 [12]

Machine reading to deal with scale problems in the making and assessing of maps

Clearly, if the IPCC is to continue producing comprehensive assessments, it has to engage in machine-reading in order to remain anchored to the wider literature. Without such an approach, it becomes harder to justify which ever-diminishing proportion of the wider literature is included

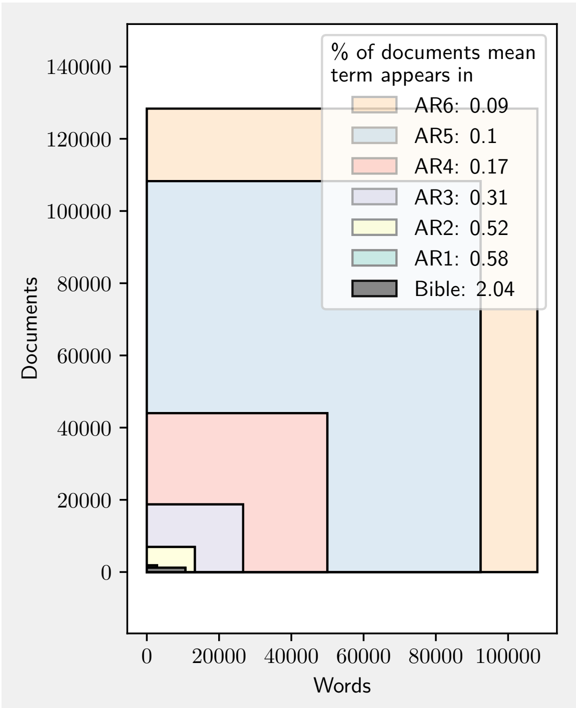


Figure 1: The volume and variety of literature on climate change has grown to unmanageable proportions. Each box represents a document-term matrix (unique documents x unique terms) of the abstracts written in each assessment period. The percentage of documents in which the average word occurs in is given in the key.

in assessments. Similarly, it becomes harder to criticise, with quantitatively evidenced claims, the outcomes of assessment processes.

Dimension reduction makes possible the description in reduced form, and with less human bias, of unmanageably large datasets

[13] [14]

This reduced form description makes comparisons more useful, when cutting the dataset.

Machine reading is a supplement to assessment-making and not free from bias; a topography is not a map

Machine reading approaches can of course not replace the task of human assessment-making. The contribution that could be made, though, is to pre-process the literature, producing a topographical map, used to navigate the literature while producing a more detailed assessment with human judgement.[[]] In fact this happens already - when IPCC authors search for literature on a topic, the results which appear on the search engine they use will be subject to algorithms based on the processing of millions of records of article text and metadata. This can be done in a much more systematic way when scientists perform directed analyses of the literature at scale.

This study's contribution. Overarching themes, structure of the literature, development, relation to IPCC

This study demonstrates how dynamic topic modelling can be used to gain an overview of an otherwise unmanageably large body of literature. This overview, or topography, describes the thematic development of the climate change literature and, in a novelly systematic way, examines how comprehensively the IPCC has been able to engage with it. In pulling together strands from text-mining, bibliometrics, and the study of science and policy, this study advances our understanding of the literature on climate change and the role of the IPCC in communicating this to policy makers.

Results

A topographical map of climate change documents shows the broad structure of climate change literature

Topics cut across both disciplinary, and working group lines - but disciplinary and working group structure remains visible in the map.

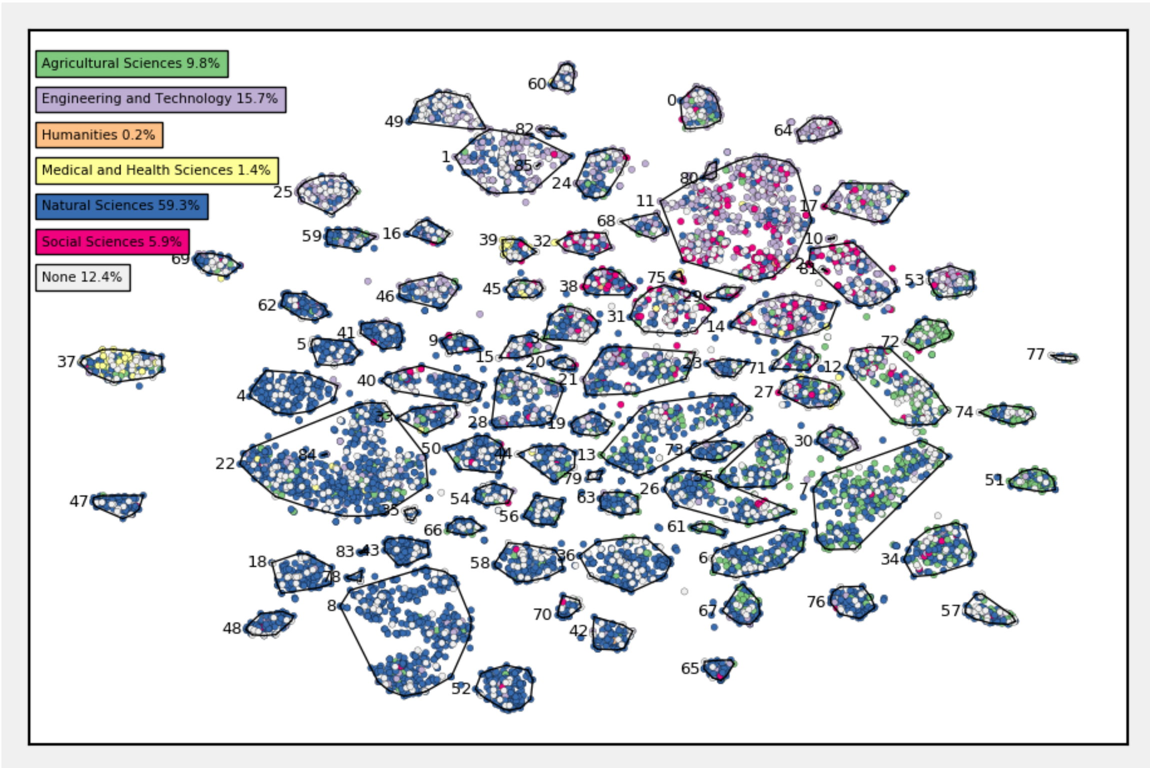


Figure 2: A map of a sample of 10,000 documents about climate change. Document positions are obtained by reducing the topic scores to two dimensions via t-SNE

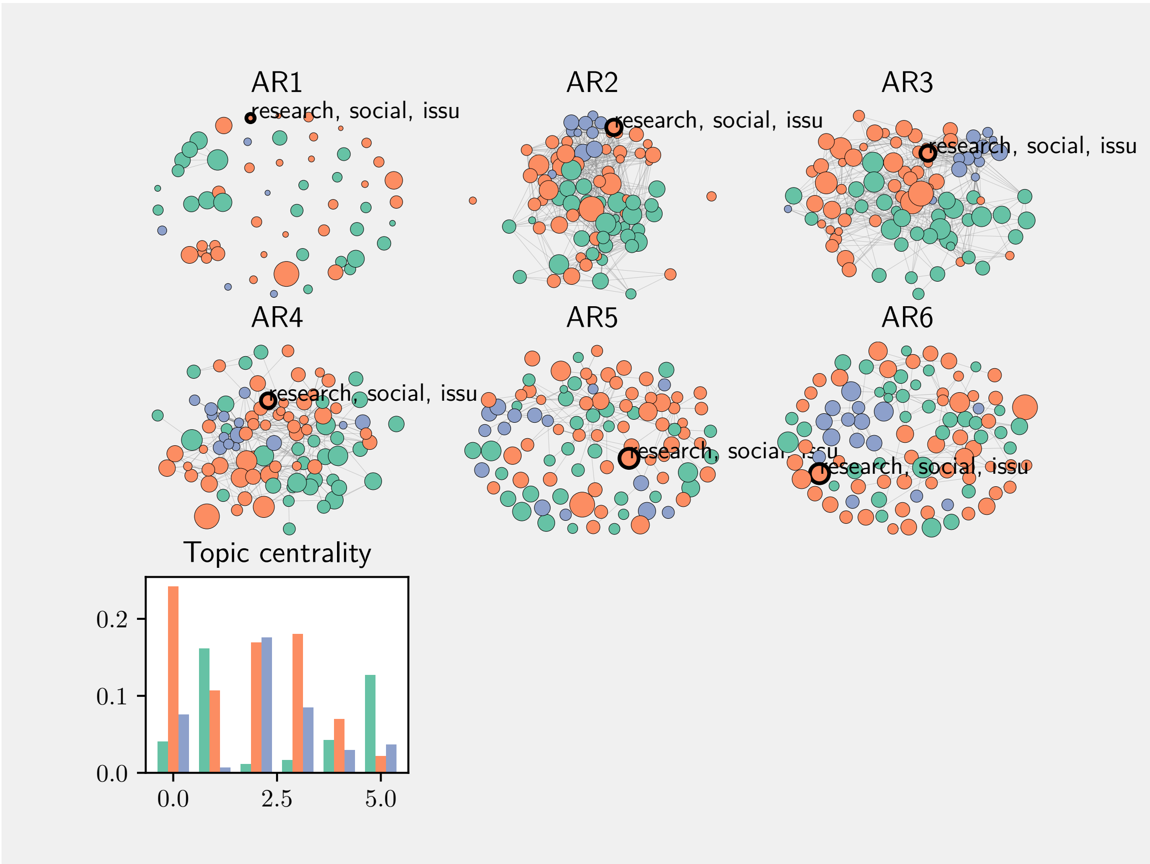


Figure 3: The development of the topic-document correlation network over IPCC assessment periods.

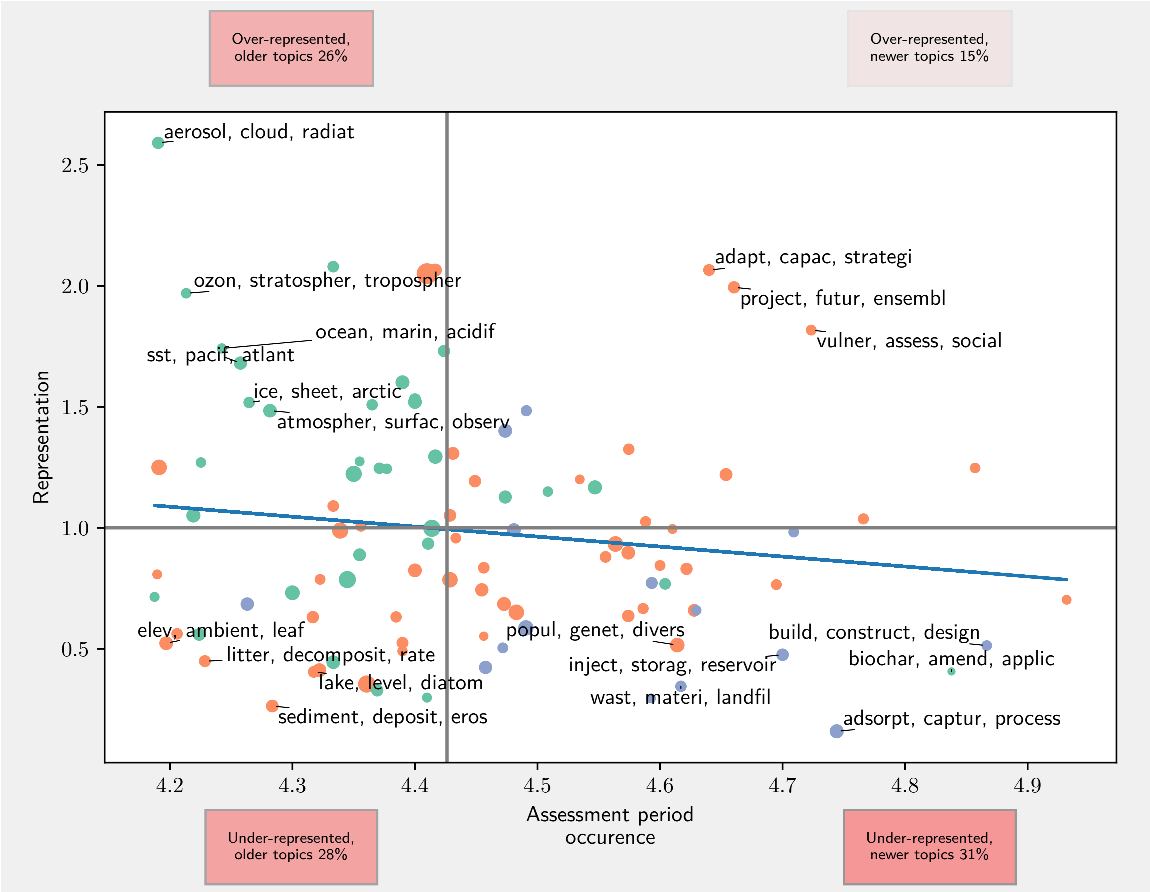


Figure 4: Representation and newness of dynamic topics

The topic-document correlation network is densest in AR2 and 3 but becomes more fragmented over time

(partly: Model less good at describing literature later on)

Working groups are clustered together [dynamics], with topics like [x] containing documents across working groups and topics like [y] important network nodes

Sustainability has been an increasingly important theme in an overarching topic about environmental sciences

(compare to biochar, which is much more recent)

Physical science topics tend to be the oldest, and the most well represented topics

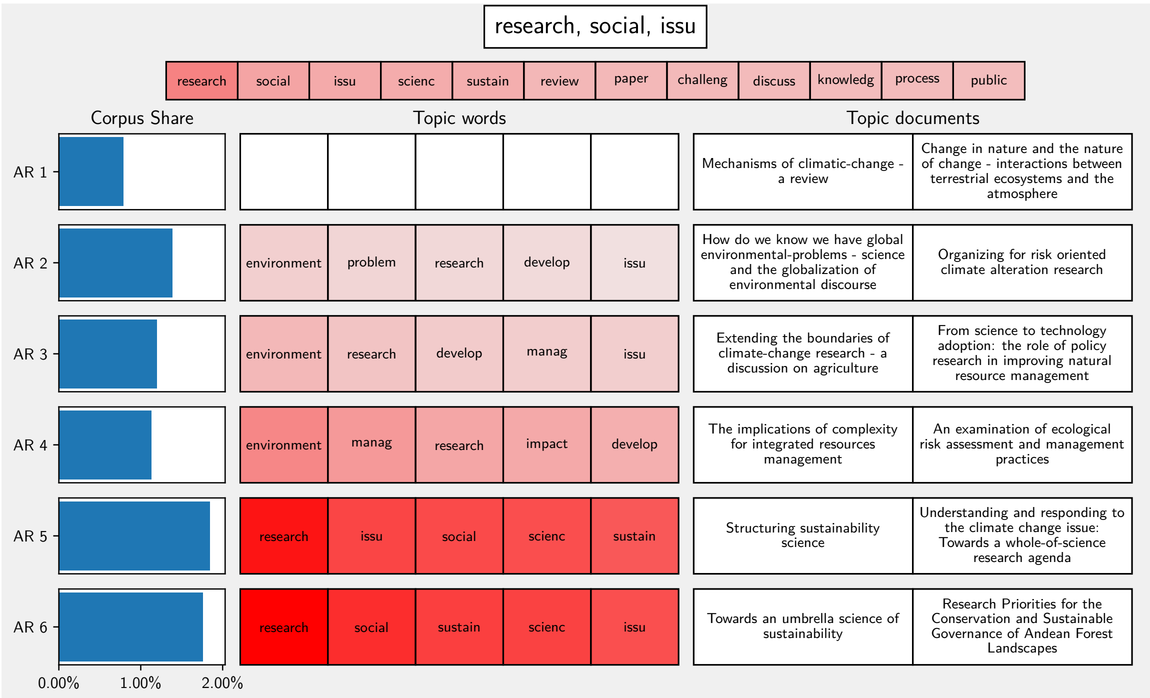


Figure 5: Word and document development of the “Research” dynamic topic

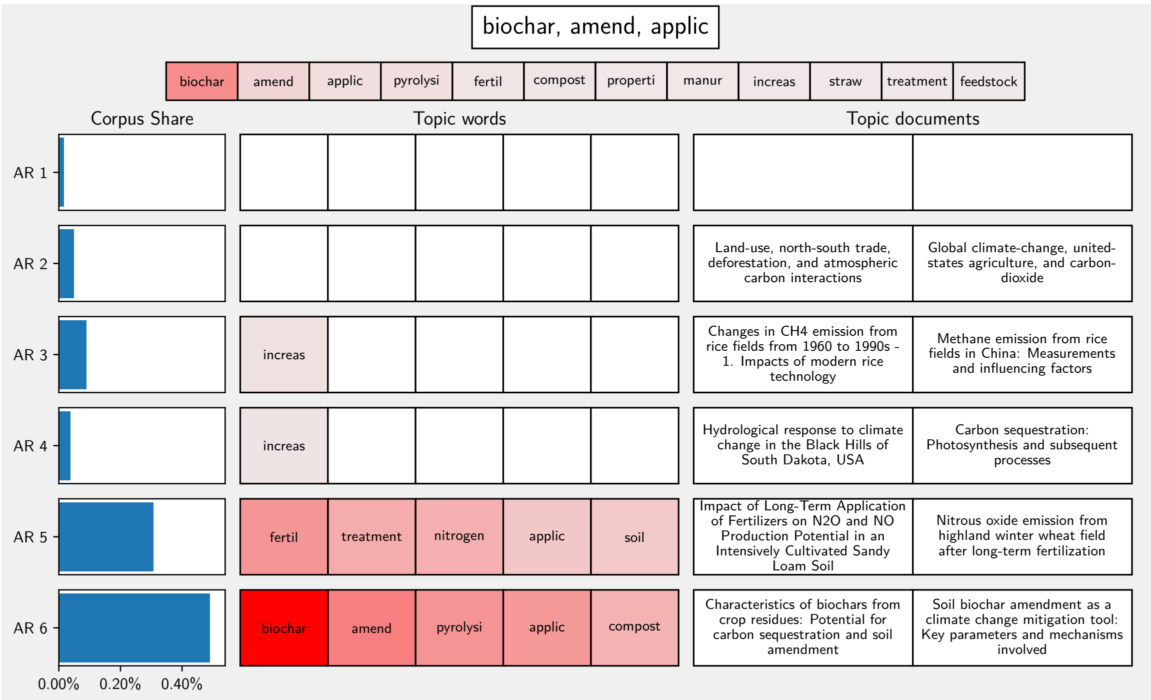
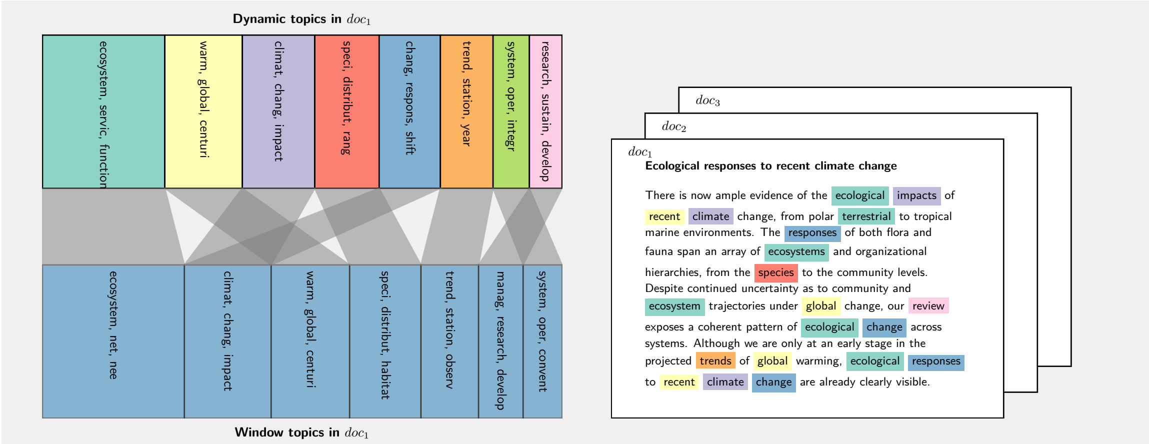


Figure 6: SI Word and document development of the “Biochar” dynamic topic



Topic / Doc	ecosystem net nee	climat chang impact	warm global centuri	speci distribut habitat
doc_1	0.034	0.021	0.019	0.017
doc_2	···	···	···	···
doc_3	···	···	···	···

Doc Topic Matrix

×

Term / Topic	chang	ecolog	global	recent
ecosystem net nee	0.18	0.39	0.01	0.03
climat chang impact	6.33	0	0.22	0.19
warm global centuri	0	0	3.72	0.31

Topic Term Matrix

≈

Term / Doc	chang	ecolog	global	recent
doc_1	4	3	2	2
doc_2	···	···	···	···
doc_3	···	···	···	···
doc_4	···	···	···	···

Doc Term Matrix

Figure 7: SI Topic make up of a single document

Adaptation and impact studies have seen a lot of growth but are well represented in IPCC reports

New topics around negative emissions and urban form are very recent and not well represented in IPCC reports.

Negative emissions in special report on 1.5, demand side chapter in AR6

Discussion

Solutions, policies and science

What do policy-makers mean when they ask for more solutions

Perfect representation is not necessarily desirable, but the skewedness should be known

There may be good reasons for a topic to be less prominent in IPCC discussions than in the wider scientific literature, and these reasons can only be understood and acted upon by humans, not by machine-reading. Nevertheless, it is desirable that assessment makers are aware of the relationship

Methods

Data

List of Figures

References

[1] IPCC. Principles governing IPCC work, 2013.

[2] Ottmar Edenhofer and Martin Kowarsch. Cartography of pathways: A new model for environmental policy assessments. *Environmental Science and Policy*, 51:56–64, 2015.

[3] Jan C. Minx, Max Callaghan, William F. Lamb, Jennifer Garard, and Ottmar Edenhofer. Learning about climate change solutions in the IPCC and beyond. *Environmental Science & Policy*, 2017.

[4] Jason Jabbour and Christian Flachsland. 40 years of global environmental assessments: A retrospective analysis. *Environmental Science and Policy*, 77(May):193–202, 2017.

[5] Andreas Bjurström and Merritt Polk. Physical and economic bias in climate change research: A scientometric study of IPCC Third Assessment Report. *Climatic Change*, 108(1):1–22, 2011.

[6] Mike Hulme. 1.5°C and climate research after the Paris Agreement. *Nature Clim. Change*, 6:222–224, 2016.

[7] Martin Kowarsch, Jason Jabbour, Christian Flachsland, Marcel T. J. Kok, Robert Watson, Peter M. Haas, Jan C. Minx, Joseph Alcamo, Jennifer Garard, Pauline Riousset, László Pintér, Cameron Langford, Yulia Yamineva, Christoph von Stechow, Jessica O'Reilly, and Ottmar Edenhofer. A road map for global environmental assessments. *Nature Climate Change*, 7(6):379–382, 2017.

[8] Robin Haunschild, Lutz Bornmann, and Werner Marx. Climate Change Research in View of Bibliometrics. *PLoS ONE*, 11(7):1–19, 2016.

[9] Michael Grieneisen and Minghua Zhang. The Current Status of Climate Change Research. *Nature Climate Change*, 1:72–73, 2011.

[10] Emily Grubert and Anne Siders. Benefits and applications of interdisciplinary digital tools for environmental meta-reviews and analyses. *Environmental Research Letters*, 11(9):093001, 2016.

[11] David M Blei, Andrew Y Ng, and Michael I Jordan. Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3:993–1022, 2003.

[12] Alex Csiszar. How lives became lists and scientific papers became data: Cataloguing authorship during the nineteenth century. *British Journal for the History of Science*, 50(1):23–60, 2017.

[13] Derek Greene and James P Cross. Exploring the Political Agenda of the European Parliament Using a Dynamic Topic Modeling Approach. pages 1–47, 2016.

[14] D D Lee and H S Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788–91, 1999.