Progress Report

Max Callaghan

1 June 2016

1 Overview

The majority of my work in 2018 has been focussed on what will be my first first-authored publication, *A Topography of Climate Change Literature.* A working draft of the introduction and some preliminary results are included in section 2, Paragraph headings are given in bold, with full paragraphs below sometimes to be flushed out. In parallel to this, and in support of the analysis, I have been further developing the tools documented in `https://github.com/mcallaghan/tmv` and described in section 3.1. These have further supported, and been improved by ideas generated from the collaborative work detailed in section 3.2. Section 3.3 elucidates other ideas for papers which will directly support my PhD.

2 A Topography of Climate Change Literature

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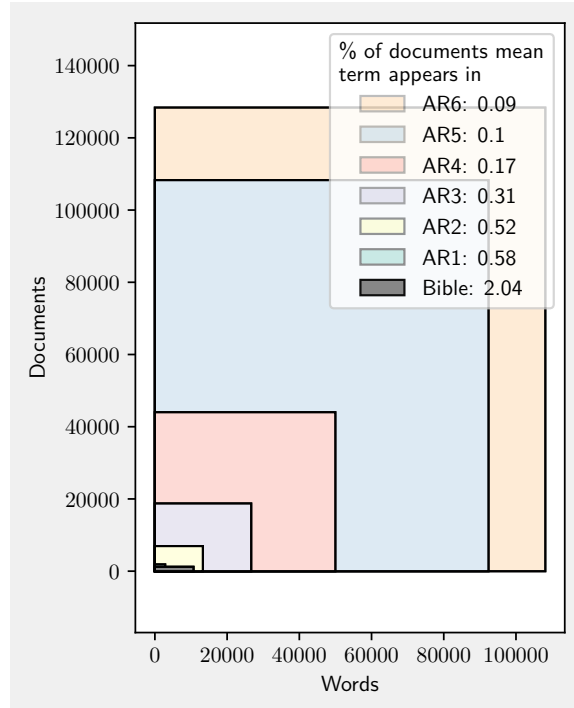
Figure 1: The volume and variety of literature on climate change has grown to unmanageable proportions. Each box represents a document-term matrix (unique documents x unique terms) of the abstracts written in each assessment period. The percentage of documents in which the average word occurs in is given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

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 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.

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)
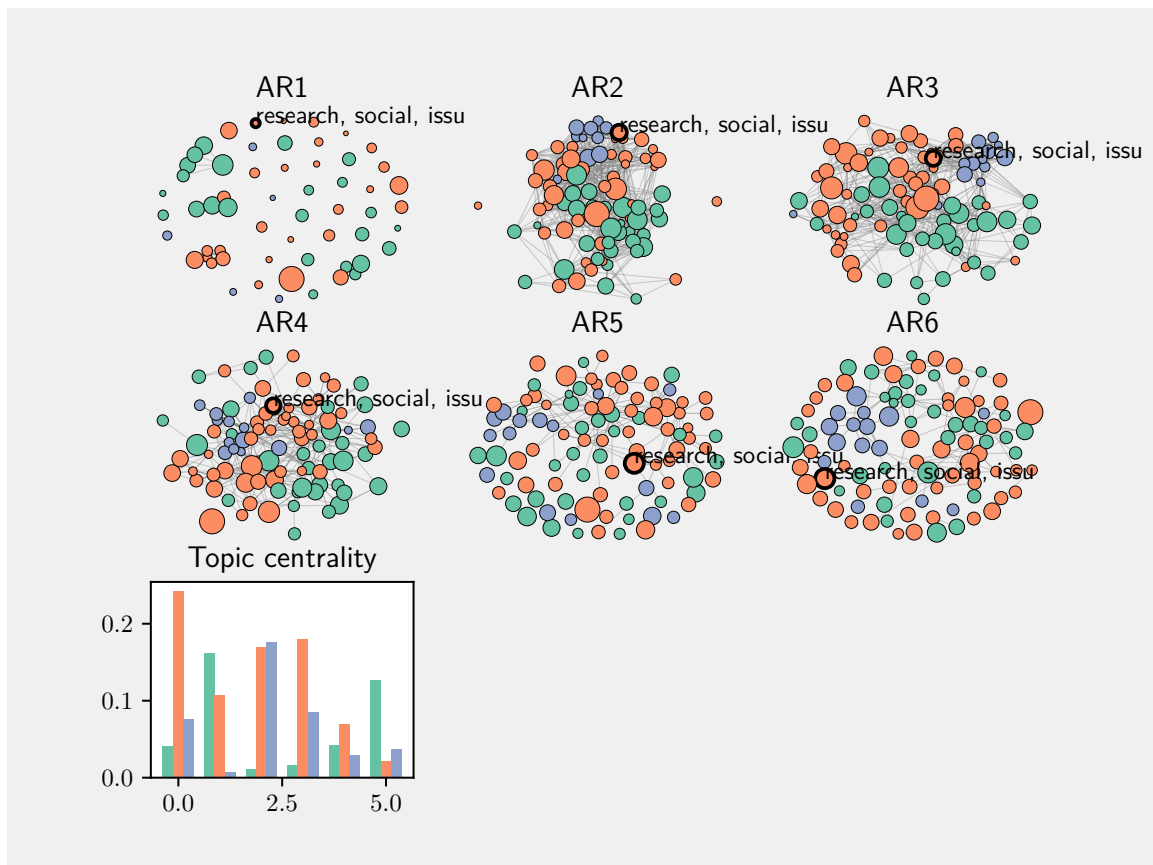
Figure 2: The development of the topic-document correlation network over IPCC assessment periods.
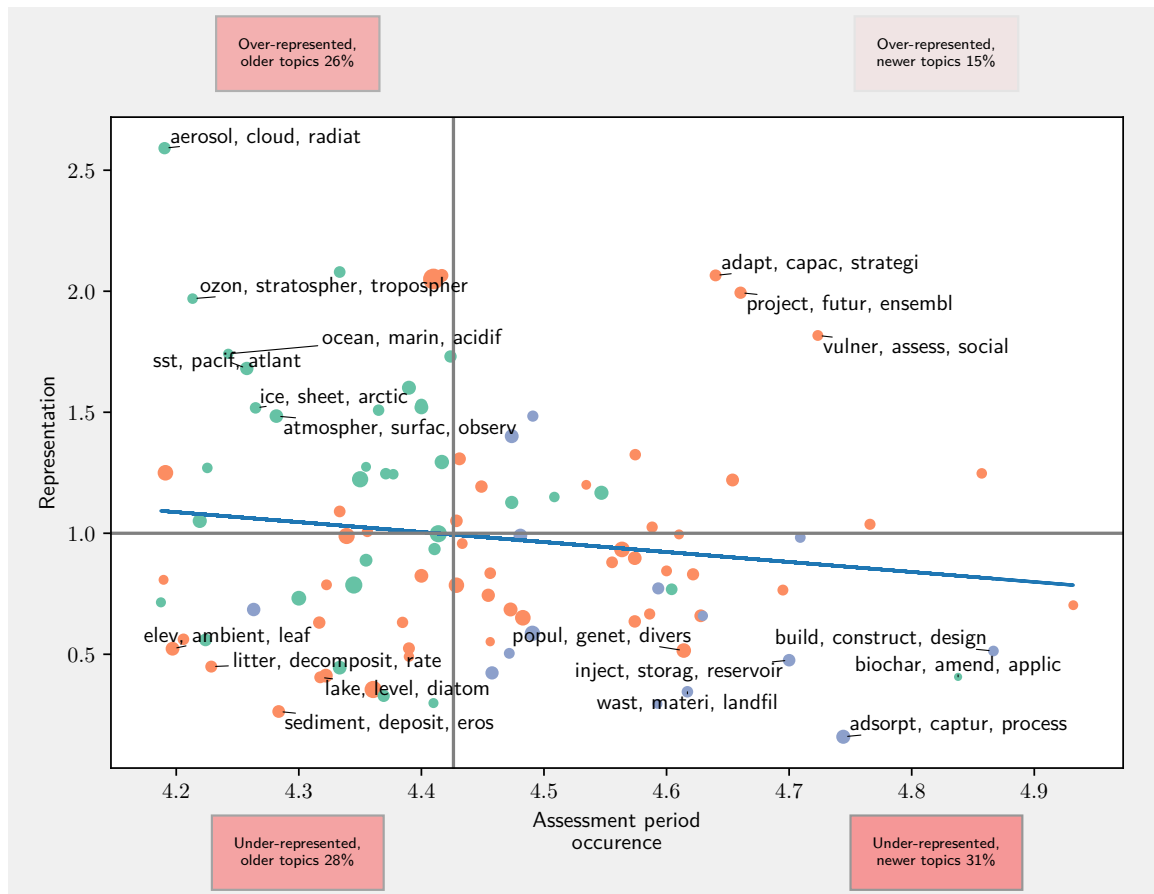
Figure 3: Representation and newness of dynamic topics

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

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

Figure 4: Word and document development of the "Research" dynamic topic

biochar, amend, applic

| biochar | amend | applic | pyrolysi | fertil | compost | properti | manur | increas | straw | treatment | feedstock |

	Corpus Share	Topic words					Topic documents	

AR 1

AR 2 — Land-use, north-south trade, deforestation, and atmospheric carbon interactions | Global climate-change, united-states agriculture, and carbon-dioxide

AR 3 — increas — Changes in CH4 emission from rice fields from 1960 to 1990s - 1. Impacts of modern rice technology | Methane emission from rice fields in China: Measurements and influencing factors
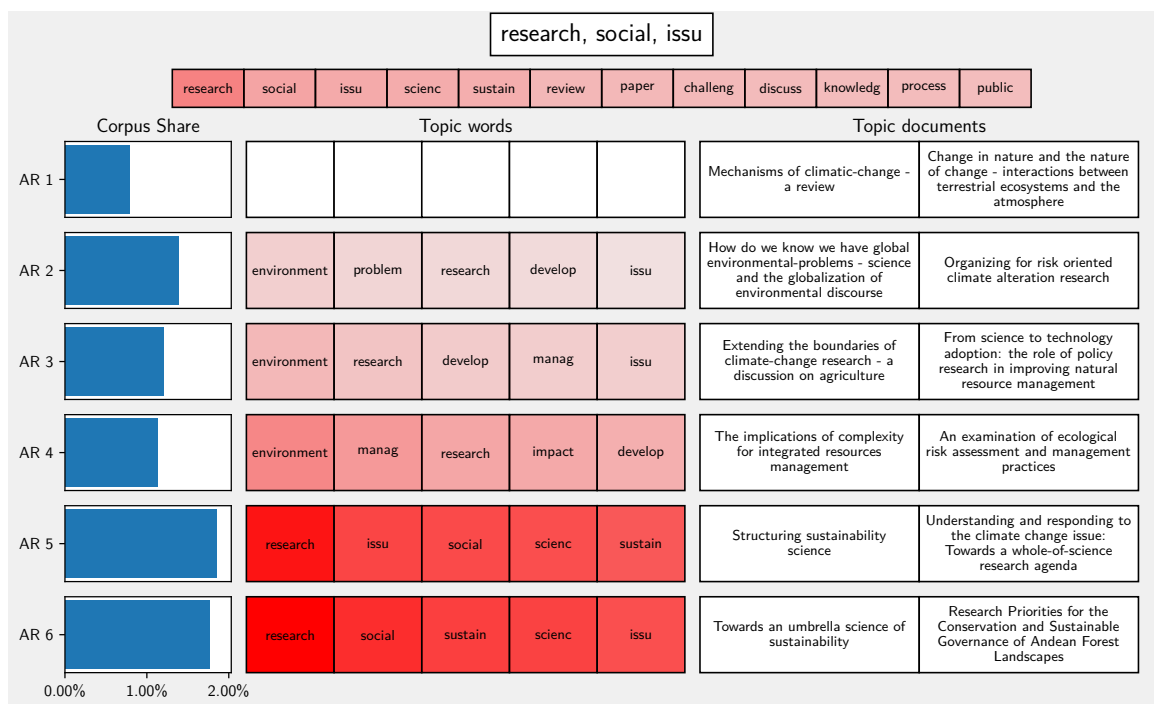
AR 4 — increas — Hydrological response to climate change in the Black Hills of South Dakota, USA | Carbon sequestration: Photosynthesis and subsequent processes

AR 5 — fertil | treatment | nitrogen | applic | soil — Impact of Long-Term Application of Fertilizers on N2O and NO Production Potential in an Intensively Cultivated Sandy Loam Soil | Nitrous oxide emission from highland winter wheat field after long-term fertilization

AR 6 — biochar | amend | pyrolysi | applic | compost — Characteristics of biochars from crop residues: Potential for carbon sequestration and soil amendment | Soil biochar amendment as a climate change mitigation tool: Key parameters and mechanisms involved

0.00% 0.20% 0.40%

Figure 5: SI Word and document development of the "Biochar" dynamic topic

8

Dynamic topics in doc_1

Window topics in doc_1

Ecological responses to recent climate change

There is now ample evidence of the ecological impacts of recent climate change, from polar terrestrial to tropical marine environments. The responses of both flora and fauna span an array of ecosystems and organizational hierarchies, from the species to the community levels. Despite continued uncertainty as to community and ecosystem trajectories under global change, our review exposes a coherent pattern of ecological change across systems. Although we are only at an early stage in the projected trends of global warming, ecological responses to recent climate change are already clearly visible.

Doc Topic Matrix

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	...	...	...	...

Topic Term 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

Doc Term Matrix

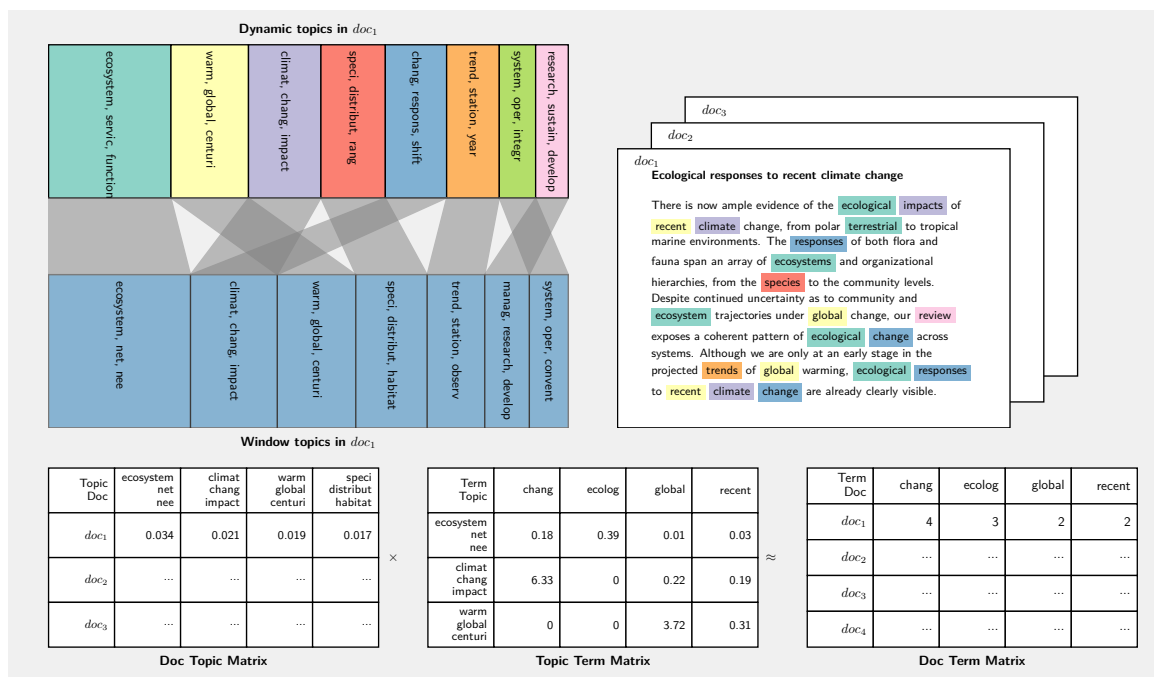| Term Doc	chang	ecolog	global	recent
doc_1	4	3	2	2
doc_2	...	...	...	...
doc_3	...	...	...	...
doc_4	...	...	...	...

Figure 6: SI Topic make up of a single document

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

3 Other Projects

3.1 Tools

The architecture of tools included in the TMV repository has expanded. The repository was built originally to view topic models interactively, then extended to manage queries and systematic reviews. The query management part has facilitated the systematic reviews on negative emissions technologies (NETs) recently published. The joining together of the query management and topic model viewing parts of the tool through an interface to run topic models on queries has allowed for wider use. Planned additions include a system to record estimates for a meta-analysis, and the incorporation of machine-learning to speed up the screening process.

3.2 Co-authored papers

3.2.1 Submitted

[15] lays out a new role for machine-reading in evidence synthesis and as such is highly relevant to my PhD. through the provision of the infrastructure to model topics, the iteration of models to choose the right topic size, the editing of the manuscript and the provision of data for the creation of figures.

[16] synthesises evidence from integrated assessment models on NETs by collecting and classifying statements about NETs. Together with first author Jérôme Hilaire, I added new functionality to the query management tool to provide for the classification of text excerpts from PDFs, and classified a portion of the statements.

3.2.2 Published

I contributed to [17] through the provision of the infrastructure to model topics, the iteration of models to choose the right topic size, the editing of the manuscript and the provision of data for the creation of figures.

[18], [19] and [20]. Form a highly ambitious systematic review on negative emissions technologies. The query management tool contributed to the efficient and accurate screening of papers across a geographically dispersed team and the management of publication data. I further contributed to the review's analysis by creating the heat bars in [19].

3.3 Paper ideas

3.3.1 Rapid Systematic Maps

This paper should lay out, and validate the methods used in [15]. The use of topic modelling to produce systematic maps is innovative: this paper should explain the methods and validate results against a human-coded dataset.

3.3.2 Quantifying the science-policy interface

In this paper, I aim, with scientometrics, to measure the properties of the knowledge pyramid, conceptualised by theorists of the science-policy interface. What kind of, and how many publications make up each layer of the pyramid, and how does information flow between layers?

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):93001, 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.

[15] William F. Lamb, Felix Creutzig, Max W. Callaghan, and Jan C. Minx. Learning about urban mitigation solutions. *Nature Clim. Change (under review)*, 2018.

[16] Jérôme Hilaire, J.C. Minx, Max W Callaghan, Jae Edmonds, Gunnar Luderer, and Joeri Rogelj. Negative emissions and international climate goals - learning from and about mitigation scenarios. *Climatic Change*, 2018.

[17] William F. Lamb, Max W. Callaghan, Felix Creutzig, Radhika Khosla, and Jan C. Minx. The literature landscape on 1.5[deg]C climate change and cities. *Current Opinion in Environmental Sustainability*, 30:26–34, 2018.

[18] Jan C Minx. How can climate policy stay on top of a growing mountain of data? | Science | The Guardian, 2018.

[19] Sabine Fuss, William F Lamb, Max W Callaghan, Jérôme Hilaire, Felix Creutzig, Thorben Amann, Tim Beringer, Wagner de Oliveira Garcia, Jens Hartmann, Tarun Khanna, Gunnar Luderer, Gregory F Nemet, Joeri Rogelj, Pete Smith, Juan Luis Vicente Vicente, Jennifer Wilcox, Maria del Mar Zamora, and Jan C Minx. Negative emissions - Part 2: Costs, potentials and side effects. *Environmental Research Letters*, accepted, 2018.

[20] Gregory F Nemet, Max W Callaghan, Felix Creuzig, Sabine Fuss, Jens Hartmann, Jérôme Hilaire, William F Lamb, Jan C Minx, Sophia Rogers, and Pete Smith. Negative emissions - Part 3: Innovation and upscaling. *Environmental Research Letters*, accepted, 2018.