

A Topography of Climate Change Research

Max Callaghan

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

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]. This role is vital...

Various researchers have attempted to analyse the IPCC’s assessment reports [5], criticised the assessment processes of the IPCC. [2] ref ref. 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].

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 [4]

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

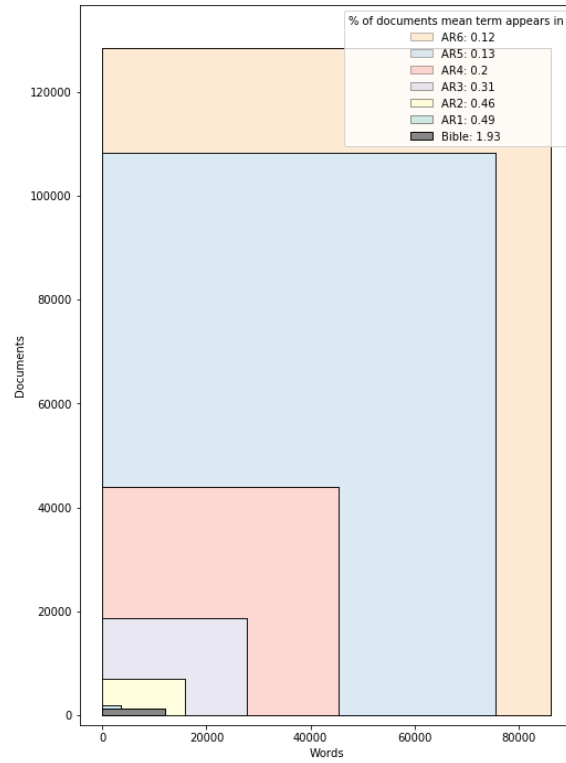


Figure 1: The volume and variety of literature on climate change has grown to unmanageable proportions. Each box represents a document-term matrix (unique documents x unique terms) of the abstracts written in each assessment period. The percentage of documents the average word occurs in is given in the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

becomes harder to justify which ever-diminishing proportion of the wider literature is included in assessments.

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.

Indeed, this same problem of scale makes attempts to criticise the IPCC as assessment-makers more difficult. These too can be strengthened by big literature approaches. + examples.

This study demonstrates how dynamic topic modelling can be used to gain an overview of an otherwise unmanageably large set 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.

Results

Increase in the frequency of the use of the term “solutions” [5],

Methods

Data

This study reproduces the query developed by [6], which is carried out on Web of Science. Though not exhaustive, it gives a good coverage of

List of Figures

- 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 the average word occurs in is given in the key. 2

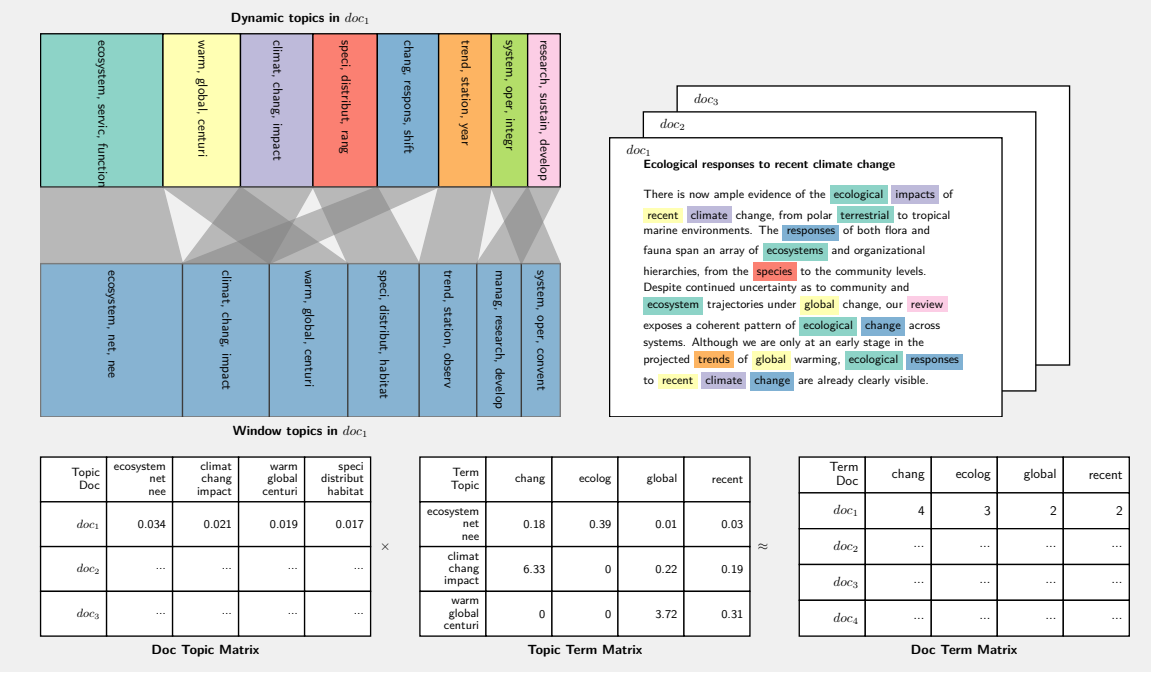


Figure 2

References

- [1] IPCC. Principles governing IPCC work, 2013.
- [2] 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.
- [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] 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.
- [5] Jason Jabbour and Christian Flachsland. 40 years of global environmental assessments: A retrospective analysis. *Environmental Science and Policy*, 77(May):193–202, 2017.
- [6] Michael Grieneisen and Minghua Zhang. The Current Status of Climate Change Research. *Nature Climate Change*, 1:72–73, 2011.