# A Topography of Climate Change Research

September 28, 2017

**To contribute to evidence-based policies on tackling climate change, the IPCC aims to comprehensively assess the relevant scientific literature (IPCC, 2013). With the size of this literature currently almost two orders of magnitude larger than at the time of the IPCC's first assessment report, this task has become impossible without the aid of machine-reading. We collect over 400,000 abstracts from Web of Science (WoS) and Scopus, and develop a topic model in order to give an overview of this unmanageably large corpus. This overview shows us the distribution and development of topics across the literature, and allows us to identify topics with greater and lesser representation in IPCC reports.**

The size of the scientific literature on climate change has expanded rapidly over the lifetime of the IPCC. While the first assessment report had around 5,500 articles to assess, nearly 5,000 new articles are now published every month, bringing the total size of the literature to close to half a million papers, (Figure 1). The increase in volume, velocity, and variety of content to be assessed has turned the task of the IPCC into a 'Big Literature' challenge (Minx, 2017). To ask questions about the literature *at scale*, we now need to apply computational techniques to the analysis of large document collections.
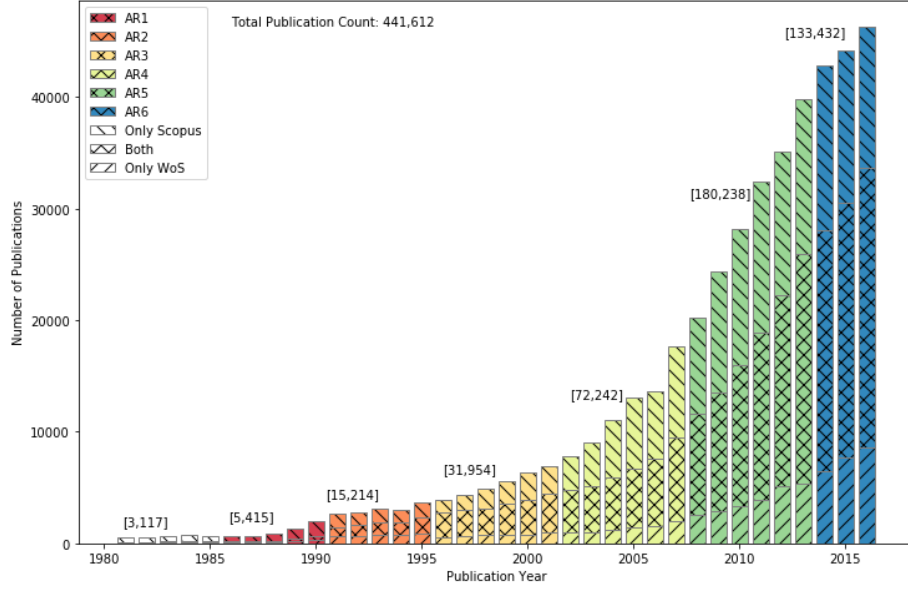
Figure 1: Growth in relevant literature in WoS and Scopus

Topic models are one such technique. A topic model learns the latent topics that structure a large corpus of documents, by leveraging the systematic co-occurrence of words across documents. Topics are distributions of words, and the topic mixture of each document explains the words observed in that document. This means that topic models can aid the understanding of large corpuses, and of the place of individual documents within them, by showing a document or corpus as a combination of 100 or so intelligible topics, rather than combinations of thousands of words.

Assessment-makers like the IPCC have been described as cartographers for policymakers (Edenhofer and Kowarsch, 2015). As such their purpose is, summarising available scientific knowledge, to describe the problem and solution space of a policy issue. The topic model presented here is a rough map of climate change research since 1985. It shows a broad outline of the topics that make up this research and how they relate to each other, and demonstrates how this has changed over time. Such a map sheds light on the terrain of knowledge about a policy issue, making an
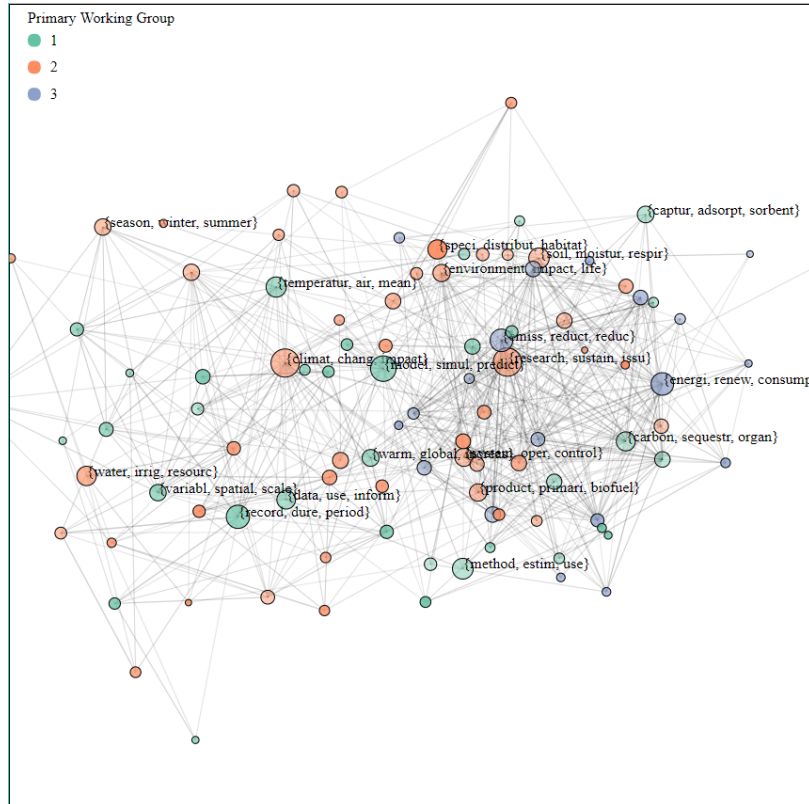
Figure 2: Topic structure of climate change literature

overview of an unmanageably large and diverse landscape possible. This overview allows both for the production of policy pathways that are well informed by science, and, as demonstrated in this paper, the assessment of the comprehensiveness with which these pathways reflect the landscape.

While topic modelling has been employed to answer specific questions about small aspects of climate literature, e.g. (e.g. Minx et al., 2017; Grubert and Siders, 2016), this is the first application of topic models to gain an overview of the entire field.

# Results

## Topic Model

### Model Structure

- Figure 2 shows the structure of the topic model with 100 topics, with each node coloured according to the IPCC working group in which the highest proportion of the topic's documents are cited.

- Topics that systematically co-occur in documents are linked, and the resulting network is displayed with links weighted according to topic correlation score. The strength of links is greater where two topics are categorised as being in the same working group. This relationship is statistically significant (see SI)

- The largest topic is formed of the words {research, sustain, issu, science, social, challeng}. This is a broad topic that focuses on meta-issues of research priorities. The documents which score most highly on this topic discuss the role of research and science in tackling sustainability problems. [E.g. Miller2014].

    - sustainability as overarching thing

    - connections of topic, wg mix of topic

    - Number of documents it occurs in

- The smallest topic contains the words {rice, paddi, field, straw, yield}. It describes the relatively discrete topic of GHG emissions from rice fields. A closely related topic (Correlation=[x]) is {methan, oxid, wetland}, as the methane emissions from rice fields are of primary concern. This is not necessarily the smallest topic in an abstract sense in climate change research.
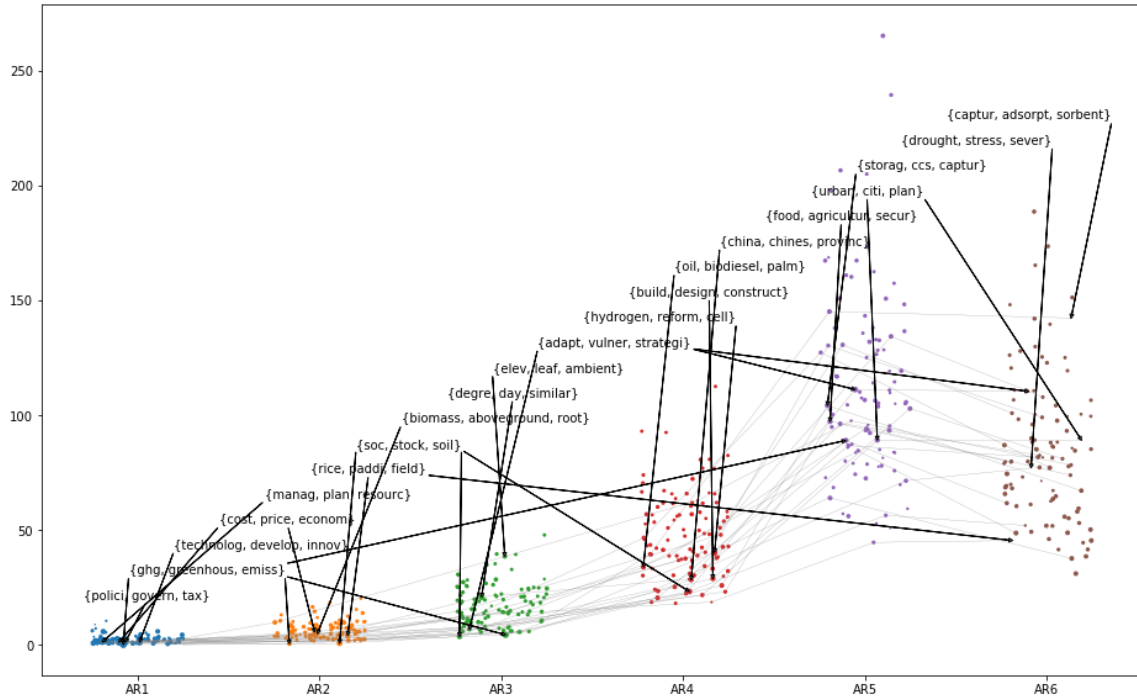
4

Figure 3: Topic growth over time. The 3 topics in each assessment period that grew by the largest amount are labelled [Example figure]

56  Rather, it is an outcome of the decision to choose 100 topics, and represents the smallest topic

57  visible at the scale implied by 100 clusters.

58  – number or documents it occurs in

59  – WG balance

## Model Dynamics

61  As well as describing the structure of a research landscape, applying a topic model can indicate areas

62  that have grown especially fast. This can give us an indicator of changing research priorities, and

63  emerging areas of study. For each topic, we sum all topic document scores in each IPCC assessment

64  period, and calculate the AR growth as the percentage increase from one period to the next.

- Figure 1 plots the growth of topics across the 6 assessment periods of the IPCC. Table highlights the five topics that show the smallest and largest growth from AR5 to AR6

    - {Captur, adsopt, sorbent} deals with the chemical process of capturing $CO_2$, either from chimneys, or directly from the air. Though this topic has been discussed for some time, e.g. [Ward1983], the stringency of climate targets implied and lack of ambitious near term mitigation implied in the Paris Agreement [cite Dependency/betting] have made this topic, as a key component of many negative emissions, all the more relevant. Here we see evidence that the policy requirements is driving technical research. [What disciplines is this research in?]

    - {urban, citi, plan} Also reflects a growing interest in cities - in relation to mitigation, decentralised, bla and impacts, urbanisation.

    - {drought, stress, sever} related to SDGs?

- For each topic, we sum the document-topic scores to calculate the size of the topic. The proportion of that sum which is accounted for by articles which we matched to the set of IPCC references can thus be seen as a measure of the coverage of each topic by IPCC reports. There is considerable variation in IPCC coverage between topics, with the extreme ends show in figure 4.

    - Is the propensity to be included in IPCC reports independent from topic scores?

    - This could be accounted for by bad matching, a larger pool to draw from (non explicitly climate papers). [Do Matching by WG, AR]

    - This is an important area for future research.

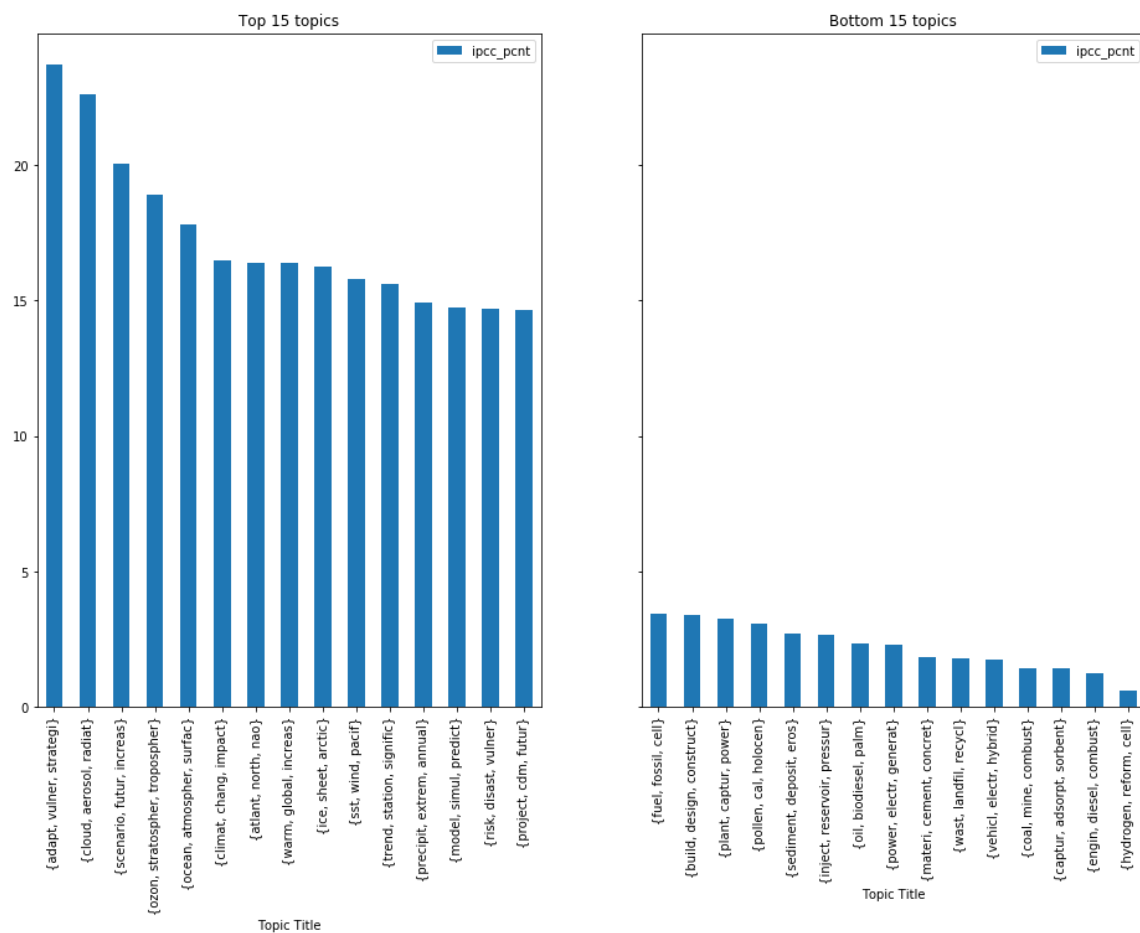- These topics are better covered in IPCC, these are less well covered 4

Figure 4: IPCC references by topic. The bars show the percentage of each topic that has been matched to an IPCC reference

| title | pchange | title | pchange |
|---|---|---|---|
| technolog, develop, innov | -39% | drought, stress, sever | 8% |
| coal, mine, combust | -40% | rice, paddi, field | 2% |
| fuel, fossil, cell | -40% | urban, citi, plan | -0% |
| hydrogen, reform, cell | -41% | adapt, vulner, strategi | -1% |
| oil, biodiesel, palm | -42% | captur, adsorpt, sorbent | -2% |

Table 1: Topics displaying the smallest (left) and largest (right) growth from AR5 to AR6

- Some interesting topic correlations are x and y; well fitting documents to both include x and y

- 

# Conclusion

- A very simple topic model provides an overview of the whole landscape.

- This allows researchers / assessment makers to identify areas that have grown recently

- Topic models aid document discovery, have the potential to contribute to more comprehensive assessments.

- AR5 seemed to have less comprehensive coverage of x topics. This was a particular issue in WG y.

- This may not be an issue, there could be good reasons for this, but these should be made transparent.

Figure 5: Focus on [biochar?] showing document with highlighted words

Figure 6: Model validation graph, showing error for different topic numbers, feature numbers

- For the next assessment report, x topics may require particular attention.

- The emerging topic on CCS resonates with a growing recognition of the importance of negative emissions and the lack of understanding about how they could fill their role. This will be of particular importance for the IPCC special report on 1.5 degrees.

# Methodology

- Topic modelling in general: reducing large matrix of documents to words to two smaller matrices of topics x words and topics x documents.

- Model selection: NMF (Lee and Seung, 1999)

- How does it work? Advantages: Simple, scalable: better results than with other solutions

- Topic model browser Chaney and Blei (2012)

- Merging with IPCC citation dataset - caveats...

- Network explanation

- Regression of network score on dummy variable for same

Figure 7: Some relation of topics to other features of dataset: e.g. most interdisciplinary journals and least, or so...

# 1 Data

- Queries: use Grieneisen and Zhang (2011), or take the best bits of Grieneisen and Zhang (2011) and Haunschild et al. (2016)?

- Sources: WoS, Scopus or both?

- Preprocessing: Remove punctuation, numbers, common, uncommon words, stemming

# List of Figures

# References

Chaney, A. and Blei, D. (2012). Visualizing Topic Models. *Icwsm*, pages 419–422.

Edenhofer, O. and Kowarsch, M. (2015). Cartography of pathways: A new model for environmental policy assessments. *Environmental Science and Policy*, 51:56–64.

132 Grieneisen, M. and Zhang, M. (2011). The Current Status of Climate Change Research. *Nature*

133 *Climate Change*, 1:72–73.

134 Grubert, E. and Siders, A. (2016). Benefits and applications of interdisciplinary digital tools for

135 environmental meta-reviews and analyses. *Environmental Research Letters*, 11(9):093001.

136 Haunschild, R., Bornmann, L., and Marx, W. (2016). Climate Change Research in View of Biblio-

137 metrics. *PLoS ONE*, 11(7):1–19.

138 IPCC (2013). Principles governing IPCC work.

139 Lee, D. D. and Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factoriza-

140 tion. *Nature*, 401(6755):788–91.

141 Minx, J. C. (2017). Big Literature in Climate Change.

142 Minx, J. C., Lamb, W. F., Callaghan, M. W., Bornmann, L., and Fuss, S. (2017). Fast growing

143 research on negative emissions. *Environmental Research Letters*, 12(3):035007.