A Topography of Climate Change Research

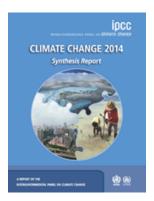
Max Callaghan



January 29, 2018

Context





- To contribute evidence-based policy-making on climate change, the IPCC aims to comprehensively assess
- These assessments should be aim to balance legitimacy, credibility and relevance (Cash and Clark, 2001)

Motivation



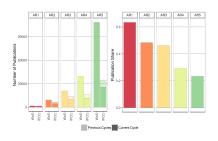


Figure: Source: Minx et al. (2017)

 Comprehensive, credible and relevant assessments become more challenging as the literature grows

To understand, and to aid, scientific assessments of climate change, we need to machine read the literature

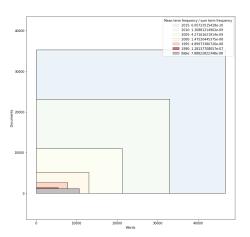
Motivation - Update 2018



Now that 2017 has ended, I have update the query to include papers from 2017 as well.

Approach - Words, words, words





- Topic modelling is a way of reducing the dimensionality of a corpus of documents
- A large matrix of documents x words is factorised by a matrix of topics x words and a matrix of topics x documents (Lee and Seung, 1999)
- Topics describe the latent structure of the document corpus



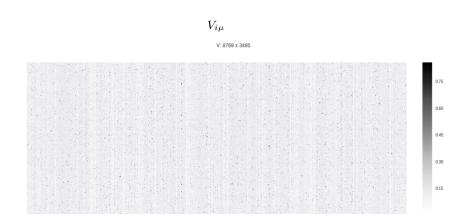


Figure: A topic model of 3495 documents on climate change from the year 2000



$$V_{i\mu} \approx (WH)_{i\mu} = \sum_{a=1}^{r} W_{ia} H_{a\mu}$$

V: 8769 x 3495

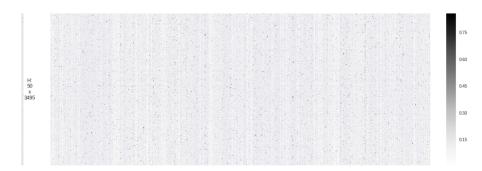


Figure: A topic model of 3495 documents on climate change from the year 2000

Max Callaghan (MCC) Climate Topography January 29, 2018 7 / 37

W: 8769 x 50

Research Questions



- What is the thematic structure of the literature on climate change, and how has this changed over the five assessment periods of the IPCC
- Which topics have had greater or less representation in the IPCC assessment reports?



The topic models above assume that the topics, and the words that make them up, are stable over time. Two approaches to better model dynamic topics:



The topic models above assume that the topics, and the words that make them up, are stable over time. Two approaches to better model dynamic topics:

 Dynamic Topic Modelling (DTM) (Blei and Lafferty, 2006) assume that a constant number of topics exists over all topic models, but allows the words in the topics to evolve from one time period to another



The topic models above assume that the topics, and the words that make them up, are stable over time. Two approaches to better model dynamic topics:

- Dynamic Topic Modelling (DTM) (Blei and Lafferty, 2006) assume that a constant number of topics exists over all topic models, but allows the words in the topics to evolve from one time period to another
- Dynamic Non-negative Matrix Factorisation (Greene and Cross, 2016) has varying numbers of topics in each window and allows for topics to emerge and/or disappear.



The topic models above assume that the topics, and the words that make them up, are stable over time. Two approaches to better model dynamic topics:

- Dynamic Topic Modelling (DTM) (Blei and Lafferty, 2006) assume that a constant number of topics exists over all topic models, but allows the words in the topics to evolve from one time period to another
- Dynamic Non-negative Matrix Factorisation (Greene and Cross, 2016) has varying numbers of topics in each window and allows for topics to emerge and/or disappear.

Where the size and variety of the literature we want to model has increased exponentially, we need an approach that allows for the emergence of new topics.



Window Topics: 1991





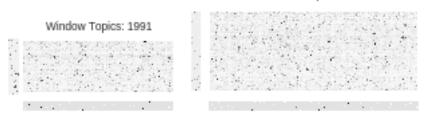
Window Topics: 1992







Window Topics: 1992



Window Topics: 1993



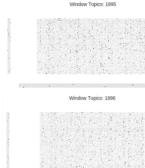


3.5	200	30.00	COUNTY CON	
	de le	1000	4	
÷.,			S. The All St.	
-	200			

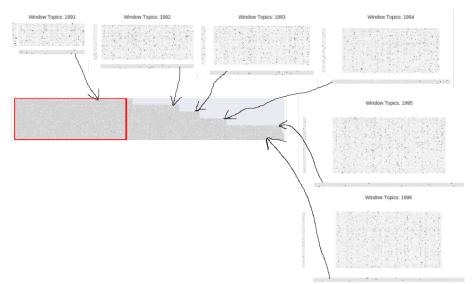




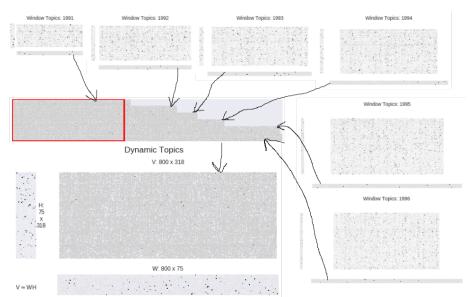












Dynamic NMF - application to climate change



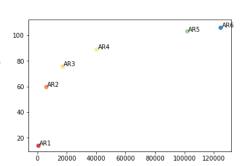
- Choosing the number of window topics is non-trivial. Data-driven approaches are limited (see below), and human selection is time consuming.
- To facilitate the description of trends over the assessment periods of the IPCC, and to minimize the number of modelling decisions, I consider each IPCC assessment period as a time window.

Dynamic NMF - application to climate change



- Choosing the number of window topics is non-trivial. Data-driven approaches are limited (see below), and human selection is time consuming.
- To facilitate the description of trends over the assessment periods of the IPCC, and to minimize the number of modelling decisions, I consider each IPCC assessment period as a time window.

 Starting from a logarithmic relationship between the number of documents and the ideal topic number, I compare 5 runs with varying numbers of topics for each window



Dynamic NMF - number of topics



Human topic number criteria

Intelligibility

Data-driven topic number criteria

Reconstruction accuracy

Dynamic NMF - number of topics

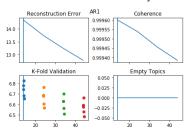


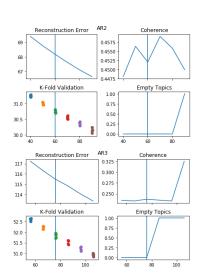
Human topic number criteria

Intelligibility

Data-driven topic number criteria

Reconstruction accuracy





Preliminary results - structure



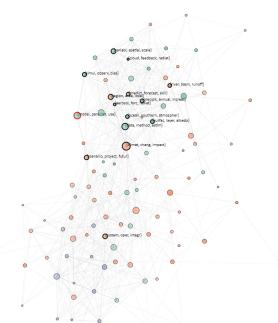


Primary Working Group

- 23
 - A network of comprehensible topics is generated with 100 topics
 - Topics can be matched to the IPCC working group from which the majority of the topic documents are referenced in
 - Topics from the same working group are significantly more likely to be correlated with each other than those which are not

Preliminary results - structure - WGI





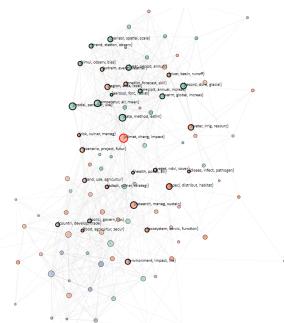
Primary Working Group

1
2

The largest topic in WGI is on models.

Preliminary results - structure - WGII





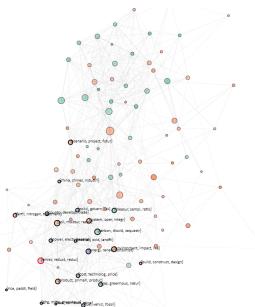
Primary Working Group

1
2

The largest primarily WGII topic is on climate change impacts

Preliminary results - structure - WGIII



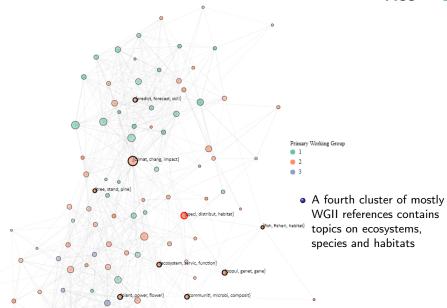


 The largest WGIII topic is on emissions reductions.

Primary Working Group

Preliminary results - structure - other clusters





Preliminary results - structure - Sustainability

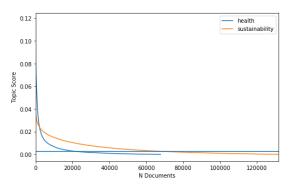




 In later assessment periods, a large meta-topic on research priorities and sustainability emerges

Preliminary results - structure - Sustainability





 The flatness of its distribution across documents indicates a topic that occurs

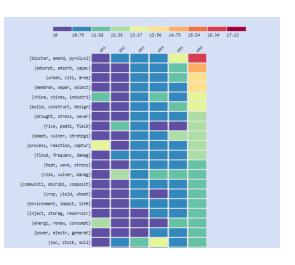
Preliminary results - structure - Sustainability



polici, govern, tax	adapt, vulner, strategi	urban, citi, area	
How Research-Prioritization Exercises Affect Conservation Policy	Methodological choices in solution-oriented adaptation research: a diagnostic framework	A review of urban ecosystem ser- vices: six key challenges for fu- ture research	
Role of hydrology and economics in water management policy under increasing uncertainty	When not every response to cli- mate change is a good one: Identifying principles for sustain- able adaptation	Lines of Tradition and Recent Approaches to Urban Ecology Focussing on Germany and the USA	
Informing food policy: balancing the evidence	Informed adaptation: Ethical considerations for adaptation researchers and decision-makers	Advancing understanding of the complex nature of urban systems	
The identification of priority policy options for UK nature conservation	Future oriented conservation: knowledge governance, uncer- tainty and learning	Sustainable urban landscapes South African perspectives or transdisciplinary possibilities	
Environmental education policy research challenges and ways re- search might cope with them	Towards an integrated agenda for adaptation research: theory, practice and policy	A conceptual framework for ad dressing complexity and unfold- ing transition dynamics when de veloping sustainable adaptatior strategies in urban water man- agement	

Preliminary results - growth

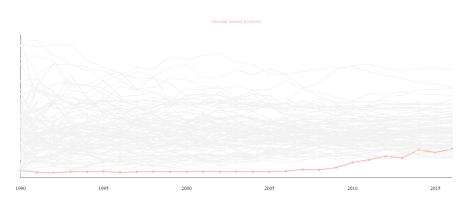




- Negative emissions related topics have shown strong growth since the end of AR5
- As have topics on cities and extreme weather events

Preliminary results - growth - biochar





• Biochar has emerged as an entirely new topic in the last 10 years

Preliminary results - gaps in coverage

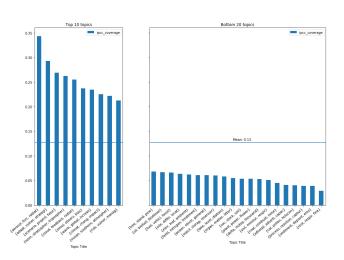


How can we get a sense of which topics are better covered in IPCC reports?

- Each document d either matches or does not match an IPCC reference
- ullet For each topic h, we can sum the scores for each category of document
- The "IPCC proportion" of each topic is the proportion of the sum of the document score accounted for by documents which match IPCC references.

Preliminary results - gaps in coverage

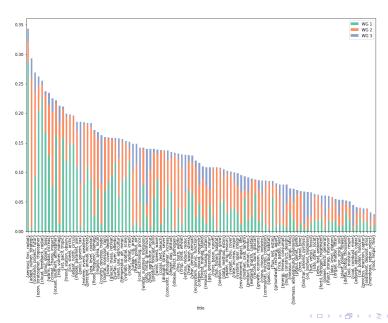




- The physical science aspects of climate change, as well topics on impacts, adaptation and scenarios are well covered by the IPCC
- Topics on specific technological solutions (particularly NETs), as well as soils, are less well covered

Preliminary results - gaps in coverage





Preliminary results - Inter-working Group topics - WG 1 and 3



IPCC Coverage	Primary WG	Topic Title	WG 1	WG 2	WG 3
0.14%	1	{winter, summer, monsoon}	0.55%	0.44%	0.01%
0.16%	2	{rainfal, monsoon, rain}	0.43%	0.56%	0.01%
0.11%	2	{season, phenolog, grow}	0.45%	0.53%	0.02%
0.16%	1	{glacier, mass, retreat}	0.60%	0.39%	0.00%
0.04%	1	{cal, pollen, holocen}	0.57%	0.42%	0.02%

Preliminary results - Inter-working Group topics - WG 1 and 3



IPCC Coverage	Primary WG	Topic Title	WG 1	WG 2	WG 3
0.13%	3	{emiss, reduct, reduc}	0.31%	0.20%	0.49%
0.09%	1	{methan, oxid, landfil}	0.61%	0.14%	0.24%
0.11%	1	{gas, greenhous, natur}	0.41%	0.24%	0.35%
0.07%	3	{fuel, vehicl, fossil}	0.23%	0.16%	0.61%
0.12%	1	{carbon, dioxid, sequestr}	0.40%	0.28%	0.32%

Preliminary results - Inter-working Group topics - WG 2 and 3 $\,$



IPCC Coverage	Primary WG	Topic Title	WG 1	WG 2	WG 3
0.19%	3	{polici, govern, tax}	0.04%	0.40%	0.56%
0.16%	3	{countri, develop, trade}	0.07%	0.43%	0.50%
0.08%	2	{build, construct, design}	0.10%	0.49%	0.41%
0.14%	3	{cost, technolog, price}	0.05%	0.30%	0.65%
0.15%	2	{urban, citi, area}	0.07%	0.62%	0.30%

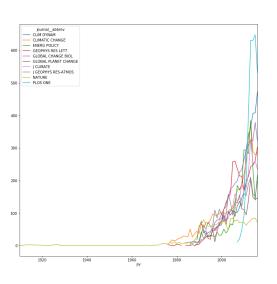
Conclusions



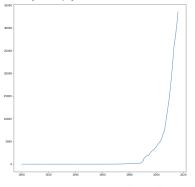
- Endogenously discovered topics make substantive sense of an unmanageable dataset of climate-relevant literature
- Topic modelling discovers over-arching topics such as that on sustainability and research priorities, as well as individual, fast growing topics such as biochar or CCS
- Matching documents to IPCC references, we can describe topics that belong to or bridge IPCC working groups
- Quantitative evidence is found to support policy makers' dissatisfaction with a lack of 'solution orientation' in IPCC reports (Kowarsch et al., 2017)
- By uncovering the structure at scale of a large collection of documents, and by enabling the discovery of documents relevant to specific topics or combinations of topics, this approach can provide for more comprehensive and salient scientific assessments.

Extra results - Journals



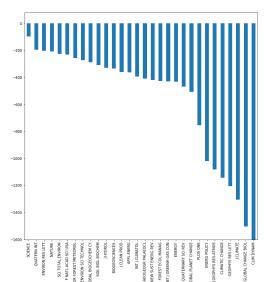


- Nature has been publishing about climate change for a long time
- Plos One has very recently overtaken all other journals
- The number of journals has risen very steeply



Extra results - Journals





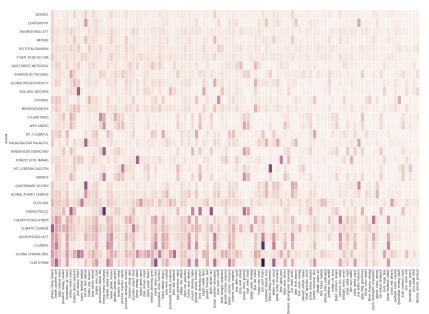
iournal abbrev

- Journal Entropy describes the distribution of topics in a journal (Hall et al., 2008)
- High values mean the journal deals with a wider range of topics

$$H(z|c,y) = -\sum_{i=1}^K \hat{p}(z_i|c,y) \ log \ \hat{p}(z_i|c,y)$$

Extra results - Journals





Frame Title



- Blei, D. M. and Lafferty, J. D. (2006). Dynamic Topic Models. *International Conference on Machine Learning*, pages 113–120.
- Cash, D. W. and Clark, W. C. (2001). From science to policy: assessing the assessment process. Social Science Research Network, (November):1–45.
- Greene, D. and Cross, J. P. (2016). Exploring the Political Agenda of the European Parliament Using a Dynamic Topic Modeling Approach. pages 1–47.
- Hall, D., Jurafsky, D., and Manning, C. D. (2008). Studying the history of ideas using topic models. EMNLP '08 Proceedings of the Conference on Empirical Methods in Natural Language Processing, (October):363–371.
- Kowarsch, M., Jabbour, J., Flachsland, C., Kok, M. T. J., Watson, R., Haas, P. M., Minx, J. C., Alcamo, J., Garard, J., Riousset, P., Pintér, L., Langford, C., Yamineva, Y., von Stechow, C., O'Reilly, J., and Edenhofer, O. (2017). A road map for global environmental assessments. Nature Climate Change, 7(6):379–382.
- Lee, D. D. and Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788–91.
- Minx, J. C., Callaghan, M., Lamb, W. F., Garard, J., and Edenhofer, O. (2017). Learning about climate change solutions in the IPCC and beyond. *Environmental Science & Policy*.