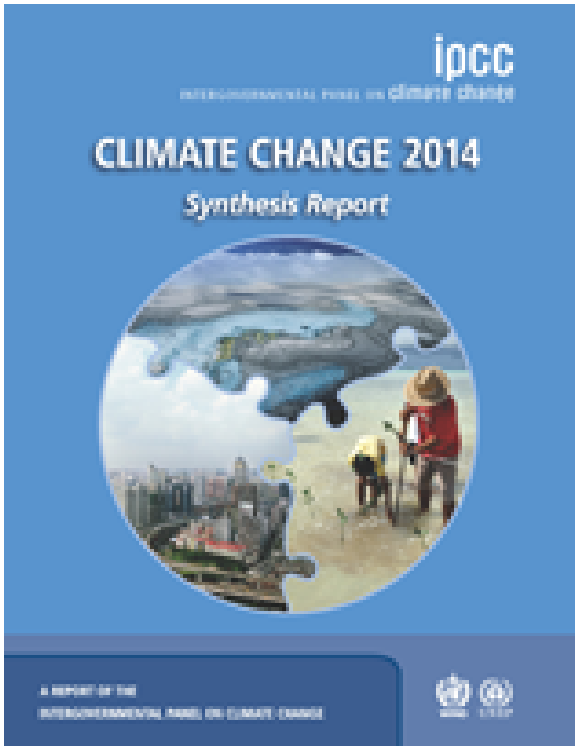
A Topography of Climate Change Research

Max Callaghan

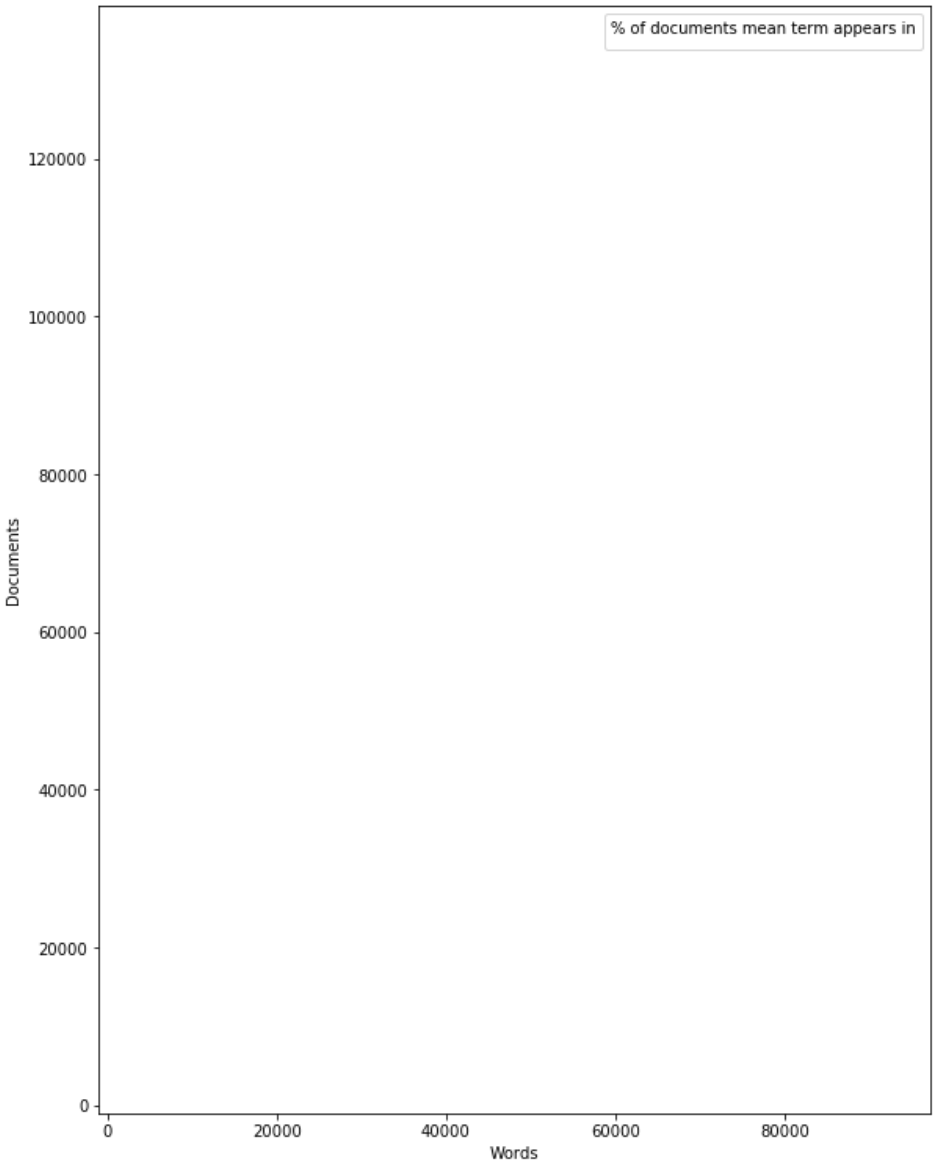
February 5, 2018

Context



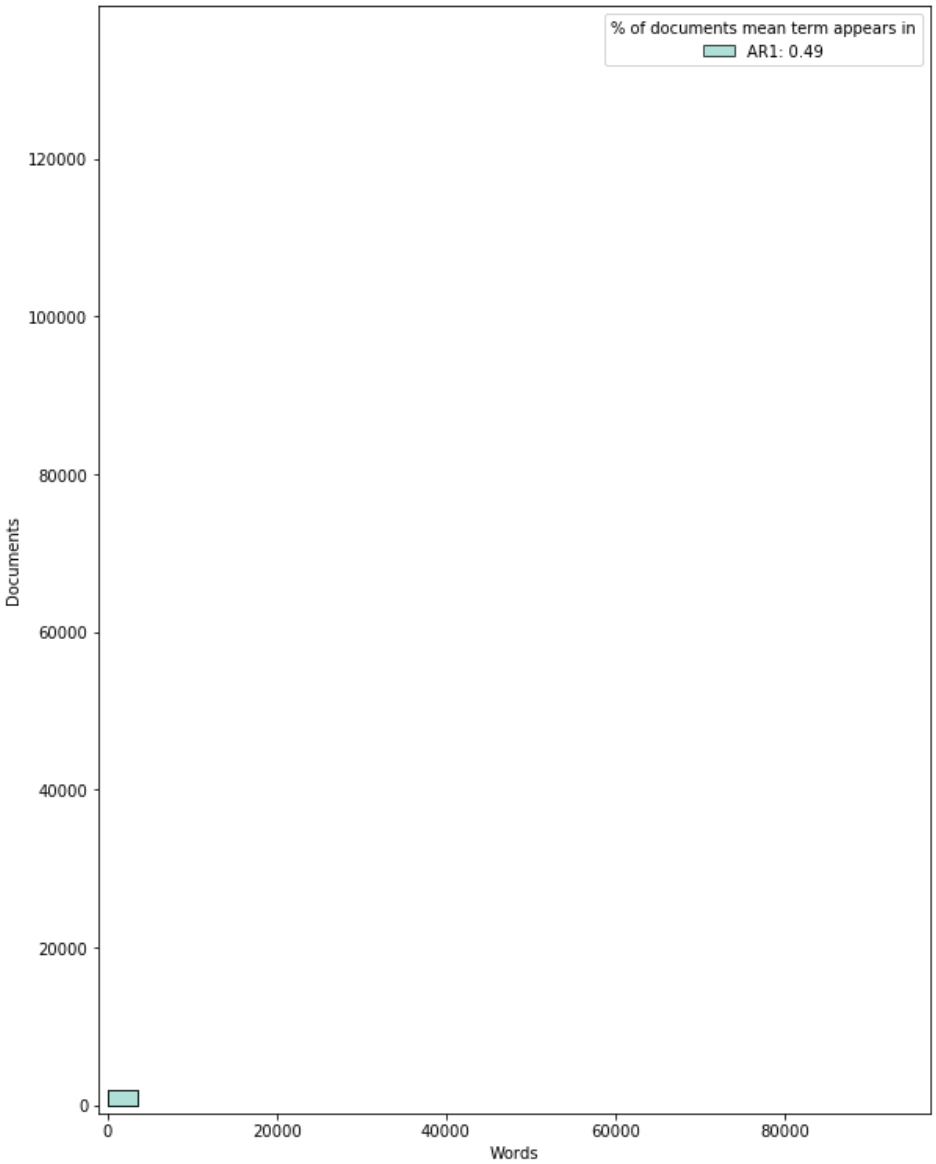
- To contribute evidence-based policy-making on climate change, the IPCC aims to *comprehensively* assess
- These assessments should be aim to balance legitimacy, credibility and relevance (Cash and Clark, 2001)

Motivation



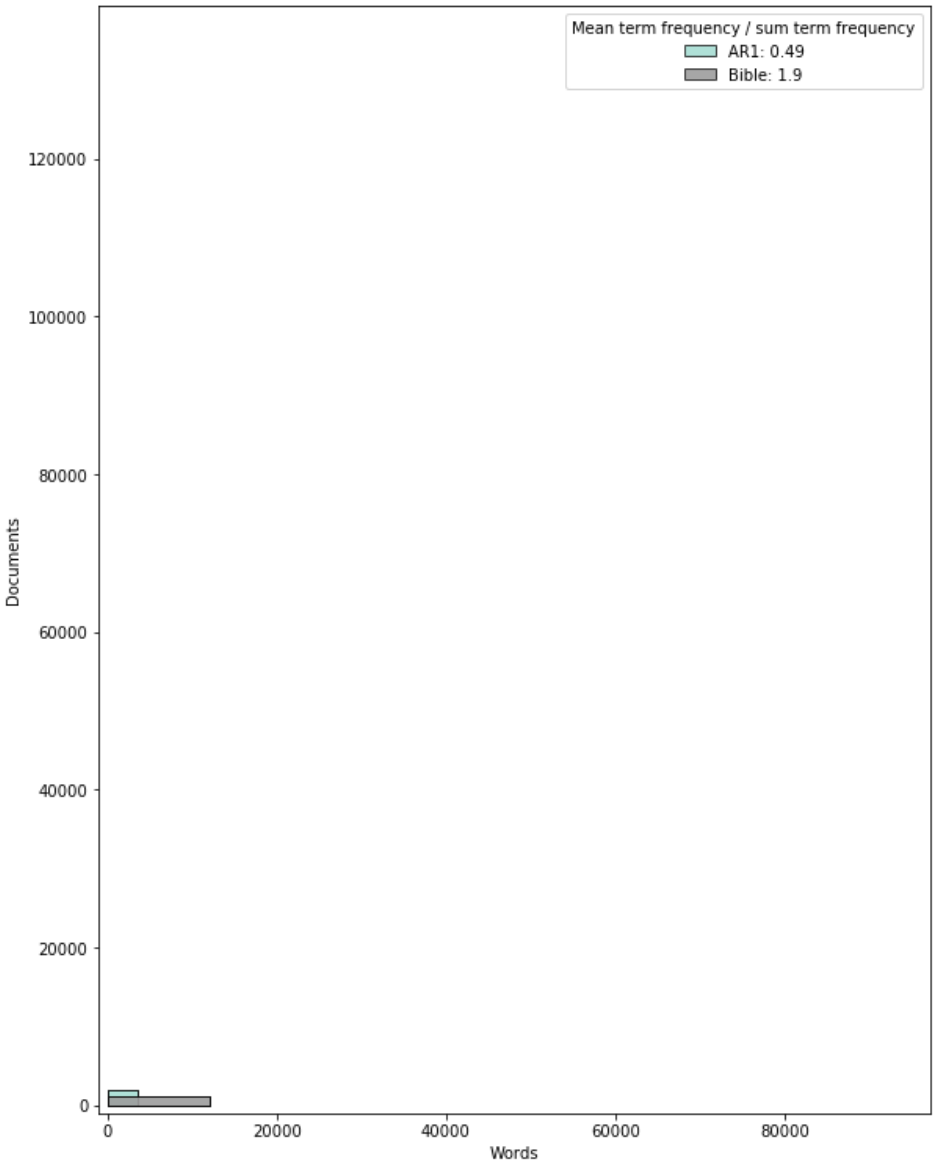
A matrix of documents x words

Motivation



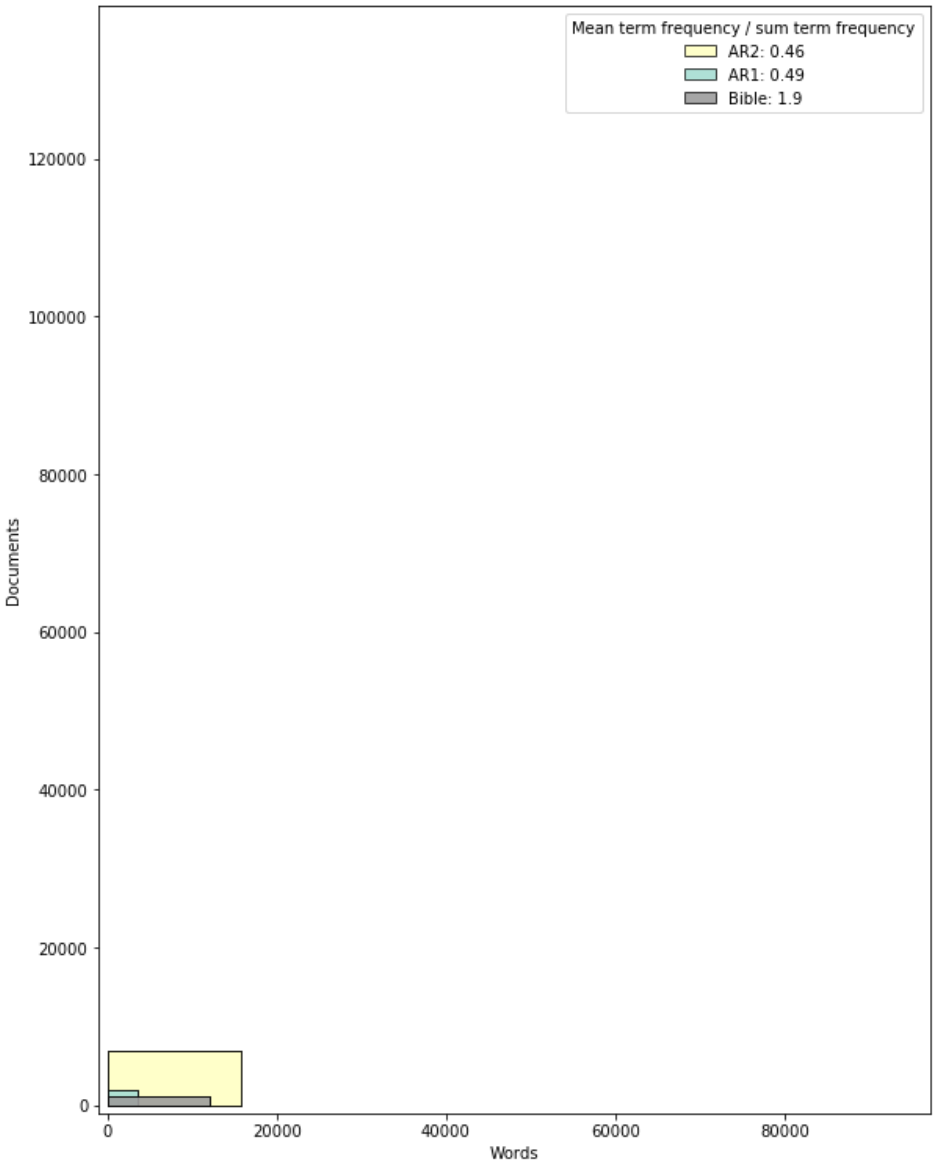
AR1: 1,848 documents x 3,528 words

Motivation



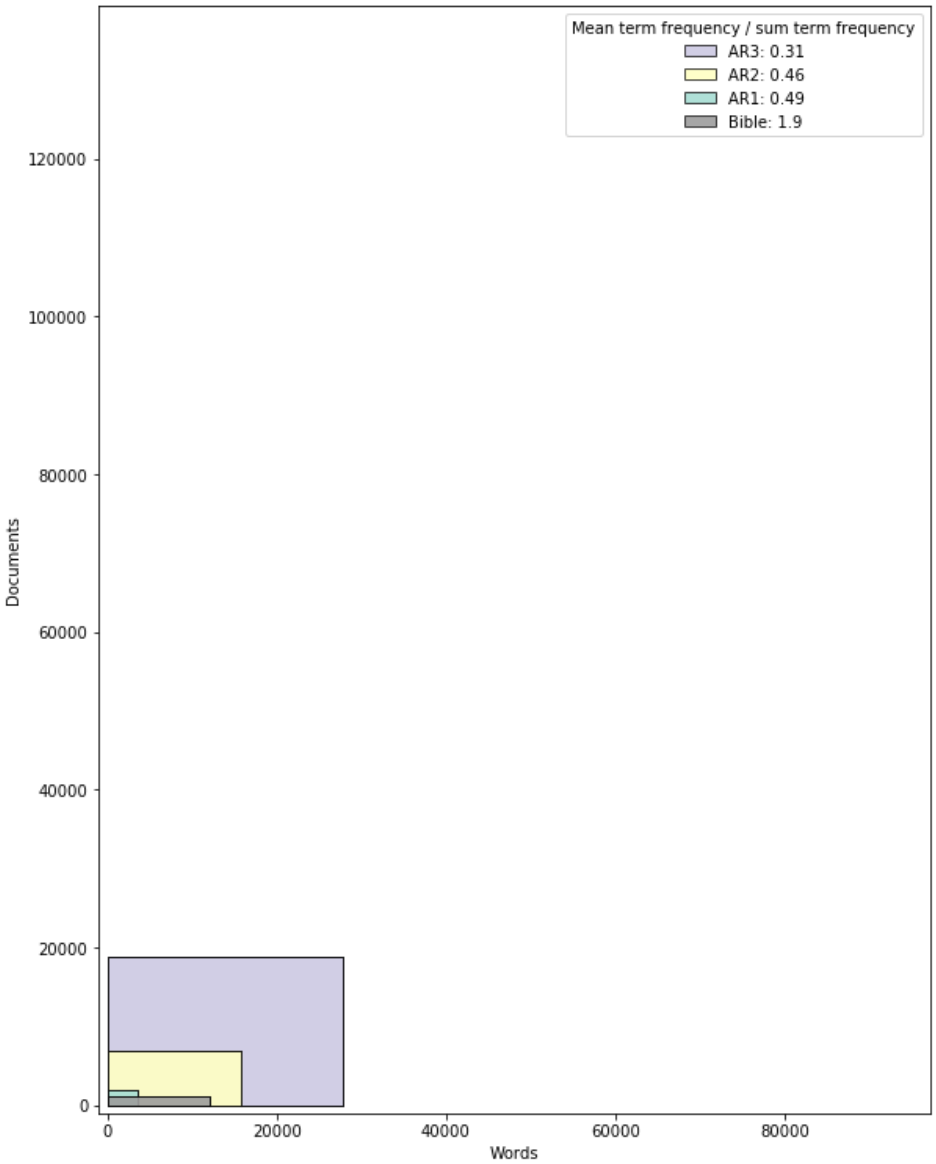
The Luther Bible: 1,189 documents (chapters) x 11,973 words

Motivation



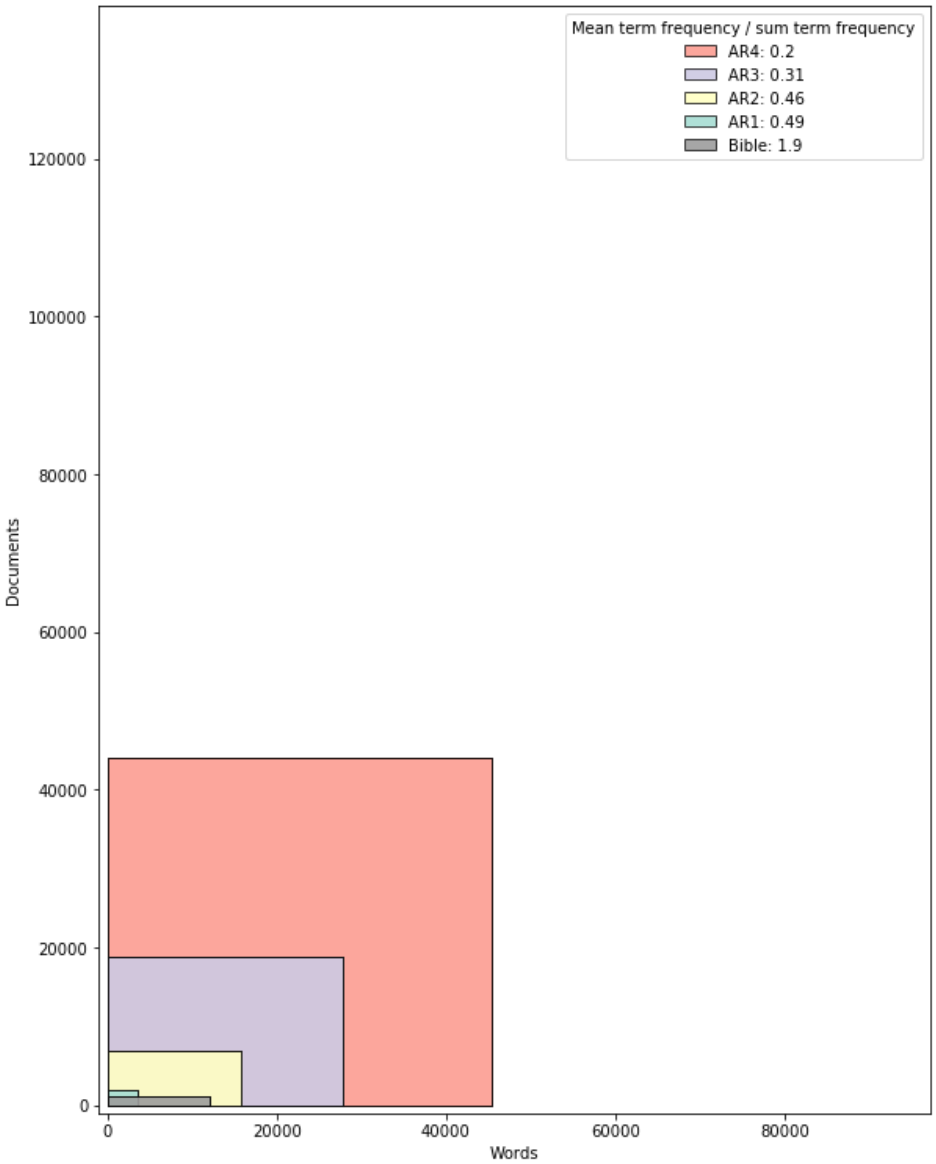
AR2: 6,941 documents x 15,781 words

Motivation



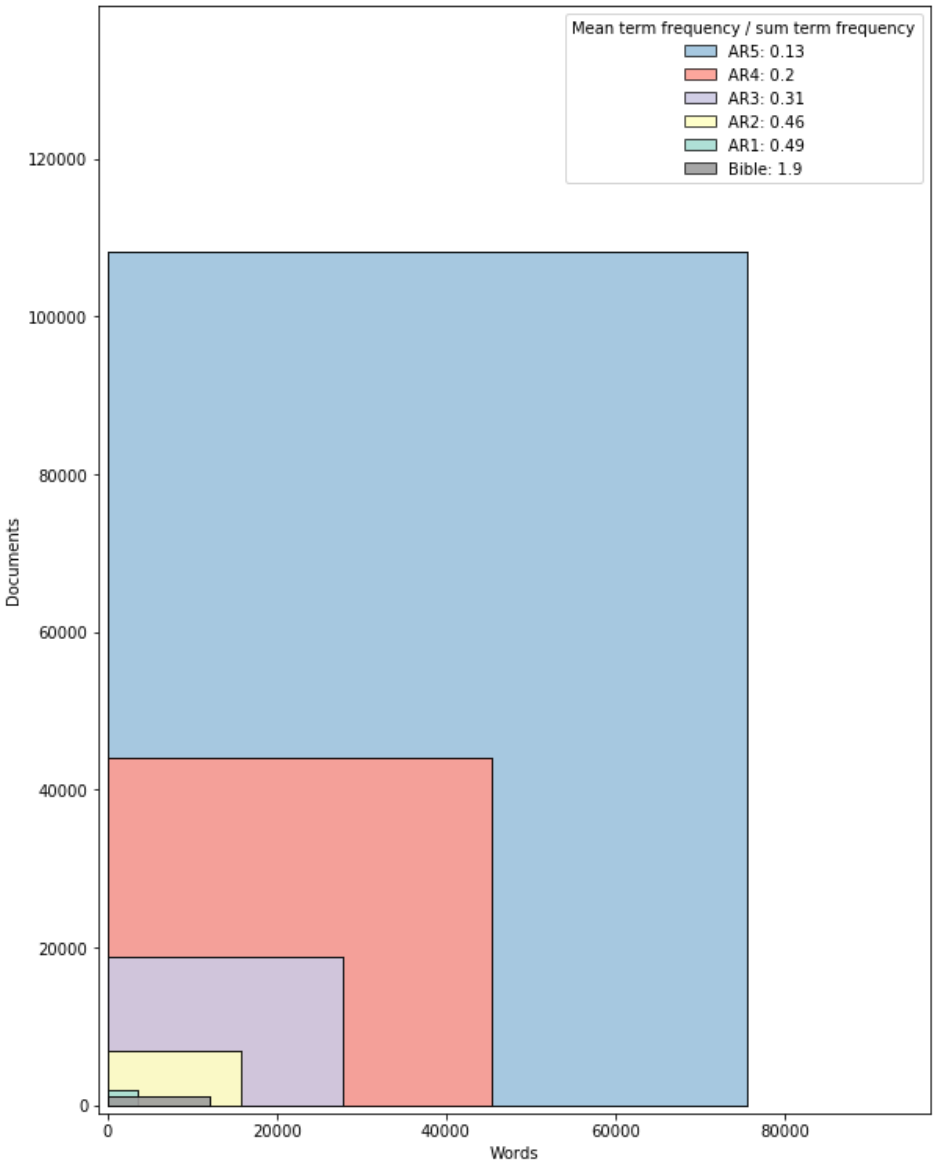
AR3: 18,728 documents x 27,730 words

Motivation



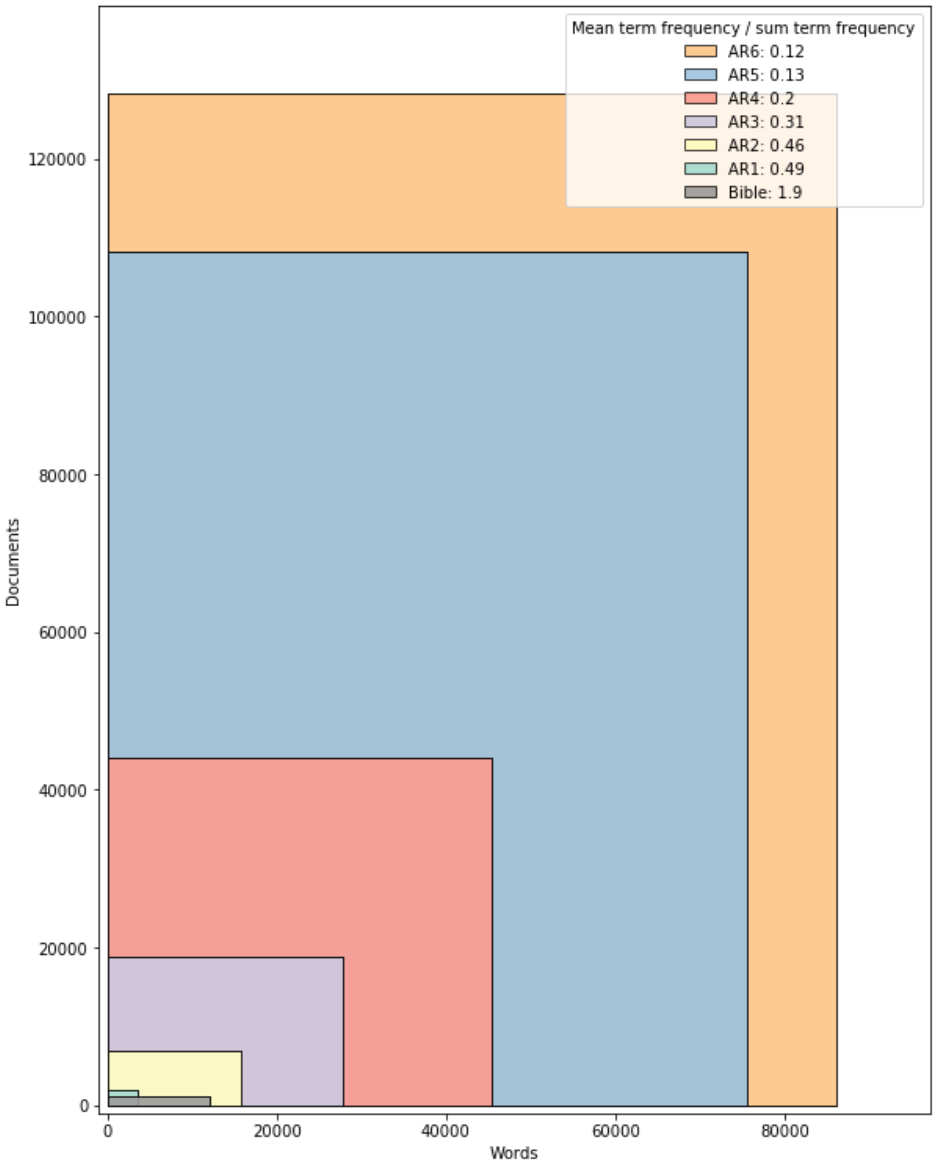
AR4: 44,000 documents x 45,388 words

Motivation



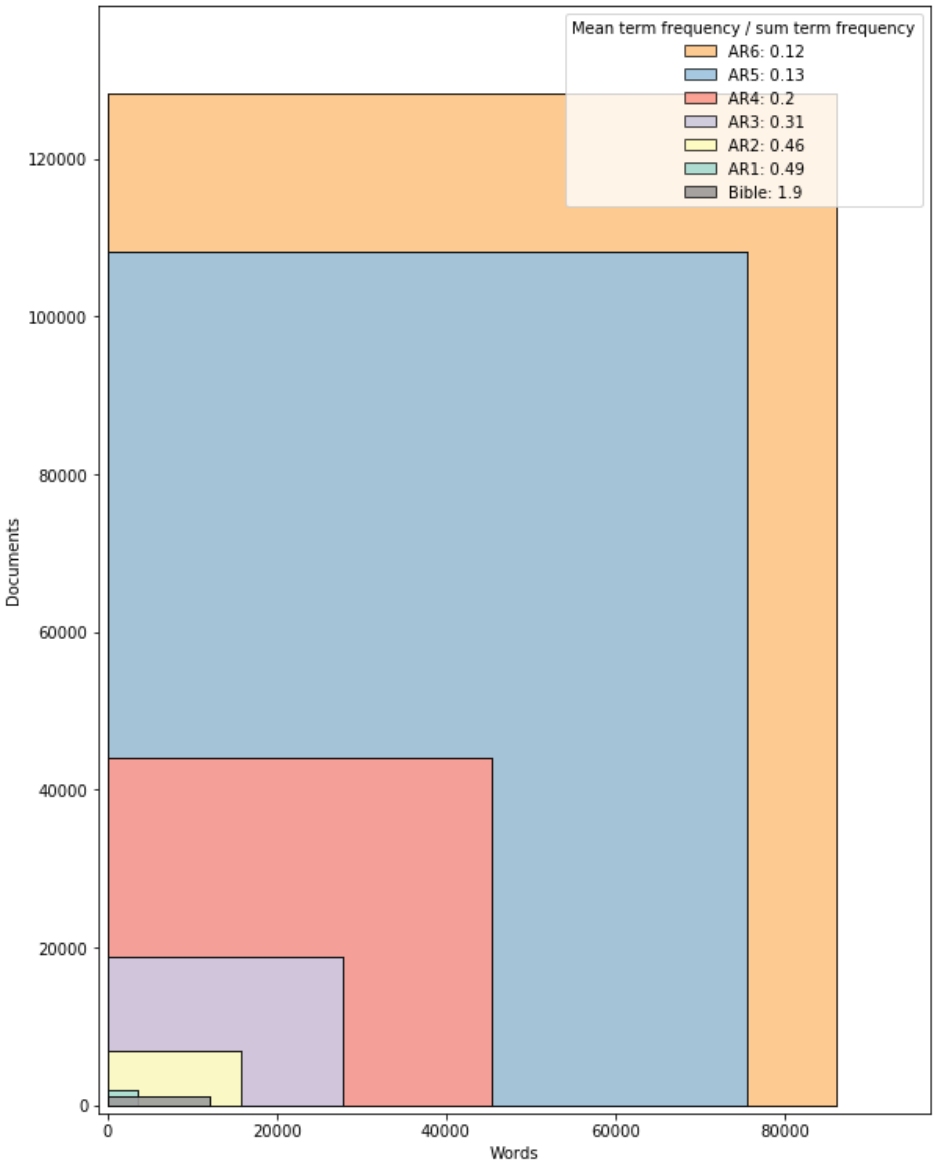
AR5: 108,277 documents x 75,553 words

Motivation



AR6: 128,357 documents x 86,149 words

Motivation

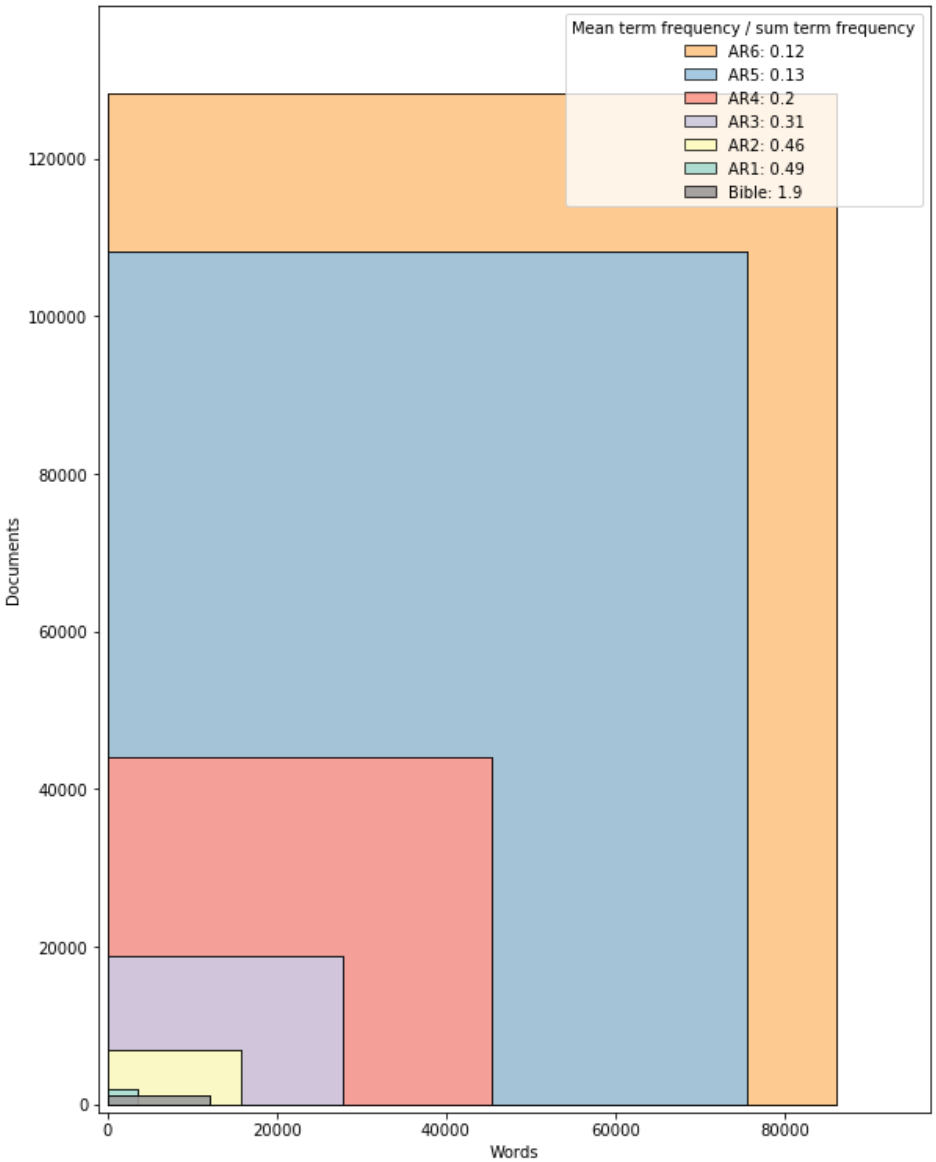


AR6: 128,357 documents x 86,149 words

- 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

Approach - Words, words, words



Topic Modelling

- 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

Approach - Words, words, words

$V_{i\mu}$ is a term frequency-inverse document frequency matrix of *stemmed* terms

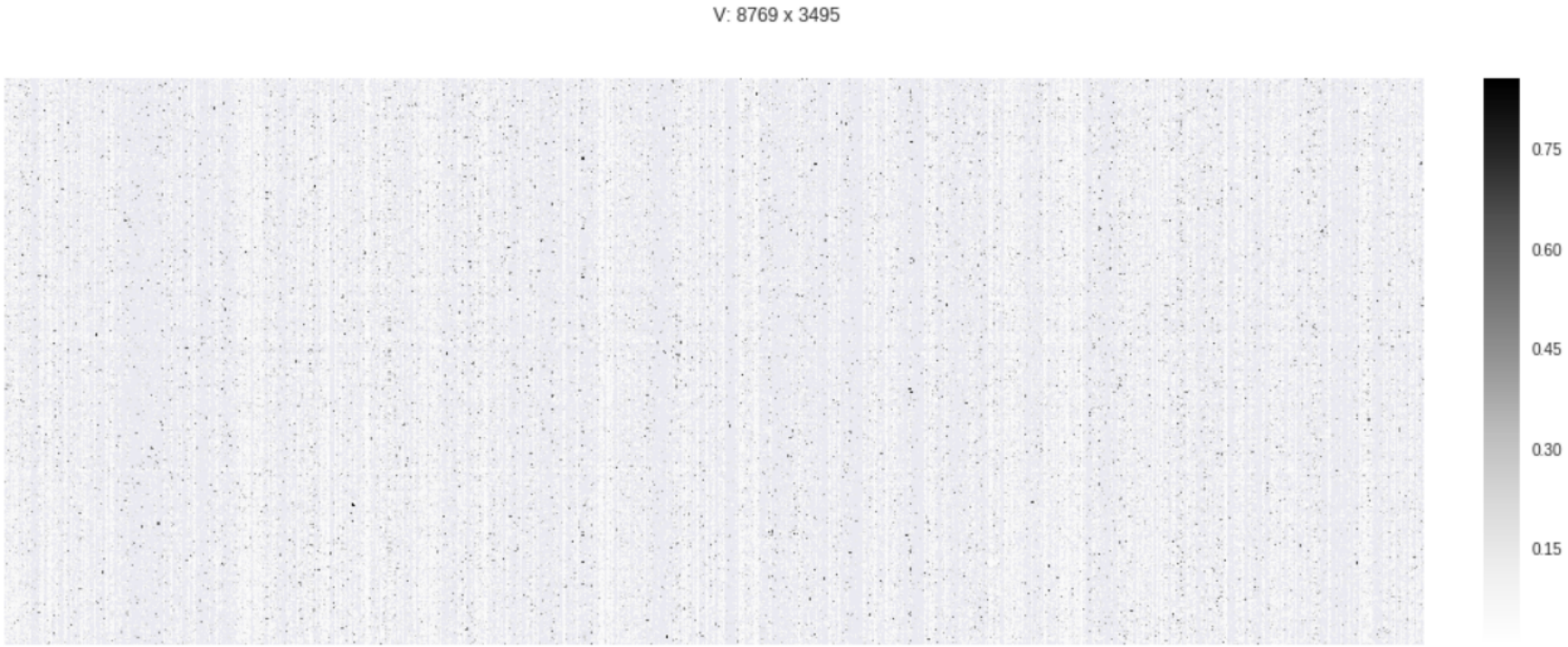


Figure: A topic model of 3495 documents on climate change from the year 2000

Approach - Words, words, words

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

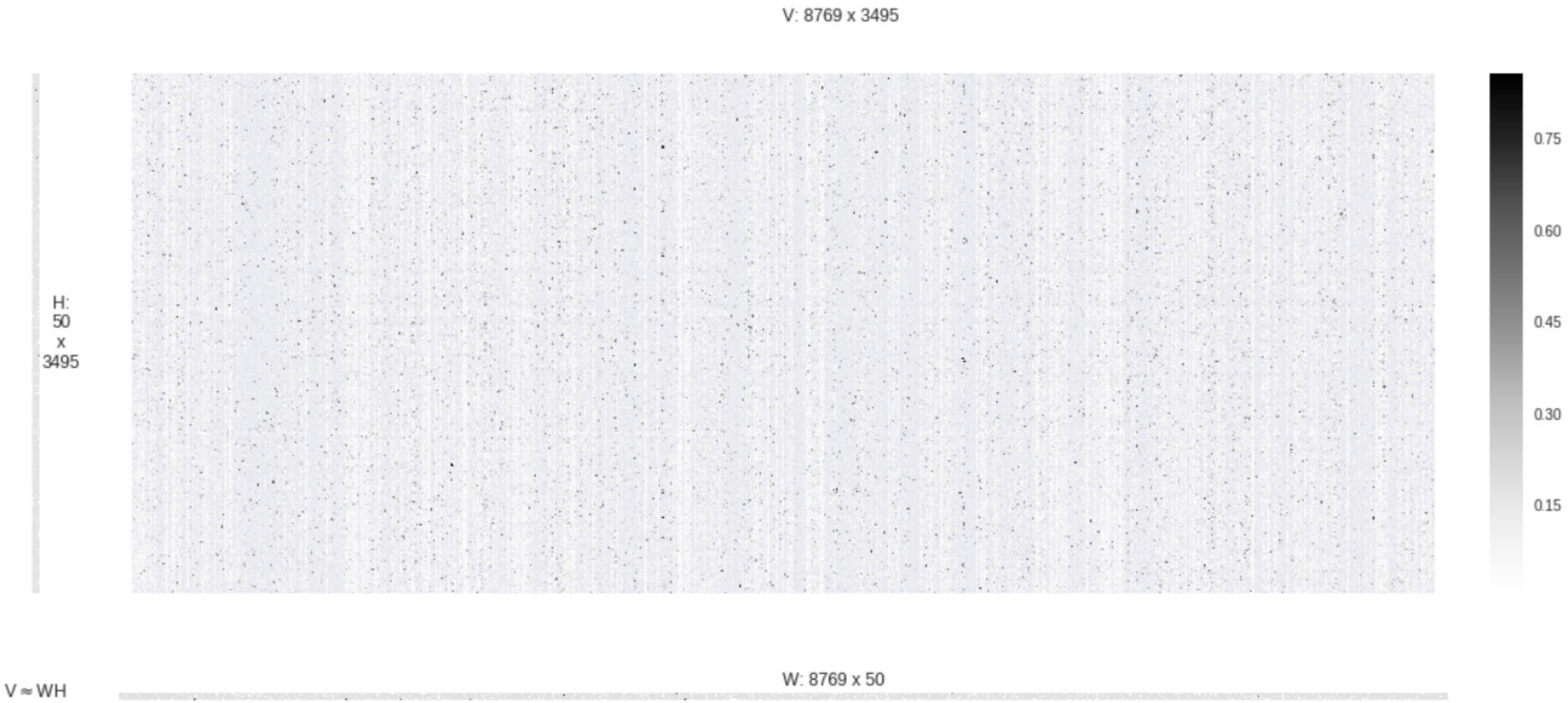


Figure: A topic model of 3495 documents on climate change from the year 2000

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

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

What can this modelled thematic structure tell us about the past and future relationship between the IPCC and scientific literature on climate change?

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

What can this modelled thematic structure tell us about the past and future relationship between the IPCC and scientific literature on climate change?

Steps

1. Download documents from Web of Science (WoS)
2. Match documents to reference lists from IPCC reports
3. Topic model stemmed document abstracts

Data - Query

MCC

(SO=(Climate Alert OR Climate Dynamics OR Climate Policy OR Climatic Change OR Global and Planetary Change OR Global Change Biology OR International Journal of Greenhouse Gas Control OR Mitigation and Adaptation Strategies for Global Change) OR TS=(((CO2 OR "carbon dioxide" OR methane OR CH4 OR "carbon cycle" OR "carbon cycles" OR "carbon cycling" OR "carbon budget*" OR "carbon flux*" OR "carbon mitigation") AND (climat*)) OR (("carbon cycle" OR "carbon cycles" OR "carbon cycling" OR "carbon budget*" OR "carbon flux*" OR "carbon mitigation") AND (atmospher*))) OR TS=("carbon emission*" OR "sequestration of carbon" OR "sequester* carbon" OR "sequestration of CO2" OR "sequester* CO2" OR "carbon tax*" OR "CO2 abatement" OR "CO2 capture" OR "CO2 storage" OR "CO2 sequester*" OR "CO2 sequestration" OR "CO2 sink*" OR "anthropogenic carbon" OR "captur* of carbon dioxide" OR "captur* of CO2" OR "climat* variability" OR "climat* dynamic*" OR "chang* in climat*" OR "climat* proxies" OR "climat* proxy" OR "climat* sensitivity" OR "climat* shift*" OR "coupled ocean-climat*" OR "early climat*" OR "future climat*" OR "past climat*" OR "shift* climat*" OR "shift in climat*") OR TS=("atmospheric carbon dioxide" OR "atmospheric CH4" OR "atmospheric CO2" OR "atmospheric methane" OR "atmospheric N2O" OR "atmospheric nitrous oxide" OR "carbon dioxide emission*" OR "carbon sink*" OR "CH4 emission*" OR "climat* policies" OR "climat* policy" OR "CO2 emission*" OR dendroclimatolog* OR ("emission* of carbon dioxide" NOT nanotube*) OR "emission* of CH4" OR "emission* of CO2" OR "emission* of methane" OR "emission* of N2O" OR "emission* of nitrous oxide" OR "historical climat*" OR IPCC OR "methane emission*" OR "N2O emission*" OR "nitrous oxide emission*") OR TS=("climat* change*" OR "global warming" OR "greenhouse effect" OR "greenhouse gas*" OR "Kyoto Protocol" OR "warming climat*" OR "cap and trade" OR "carbon capture" OR "carbon footprint*" OR "carbon neutral" OR "carbon offset" OR "carbon sequestration" OR "carbon storage" OR "carbon trad*" OR "changing climat*" OR "climat* warming")) NOT PY=2018

- (Haunschild et al., 2016)
- 309,697 documents

Caveats

- Not perfect query
- WoS not all peer-reviewed literature
- Missing grey literature
- Missing relevant literature not directly about climate change

Methodology - Dynamic Topic Modelling

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:

Methodology - Dynamic Topic Modelling

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

Methodology - Dynamic Topic Modelling

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.

Methodology - Dynamic Topic Modelling

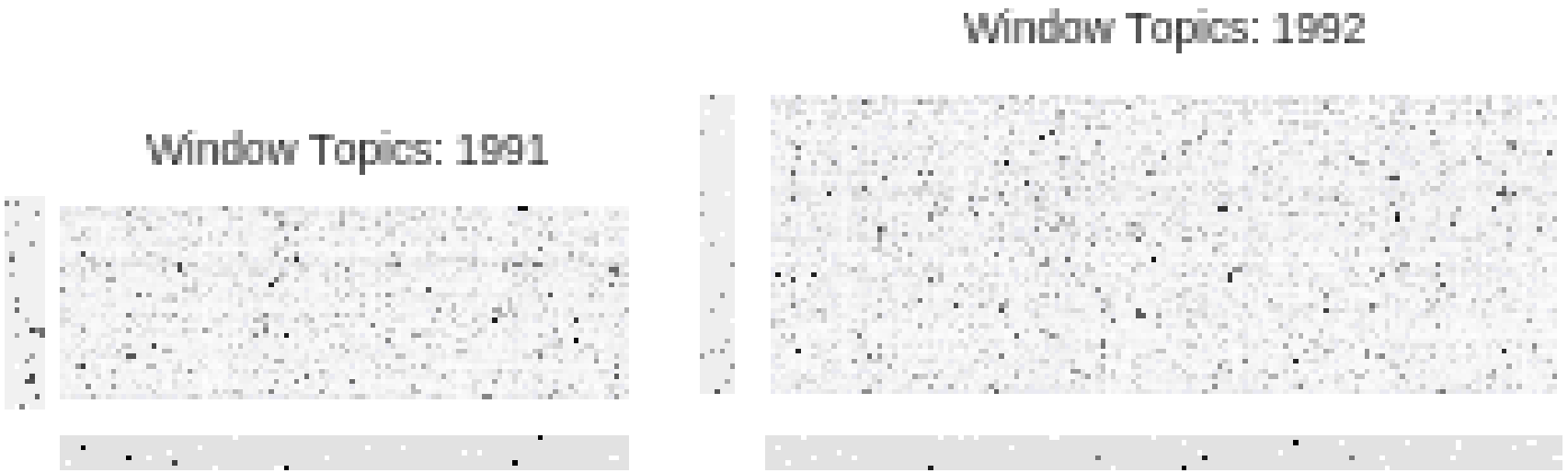
The topic models above assume that the topics, and the words that make them up, are stable over time. Two approaches to better model dynamic topics:

- 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.

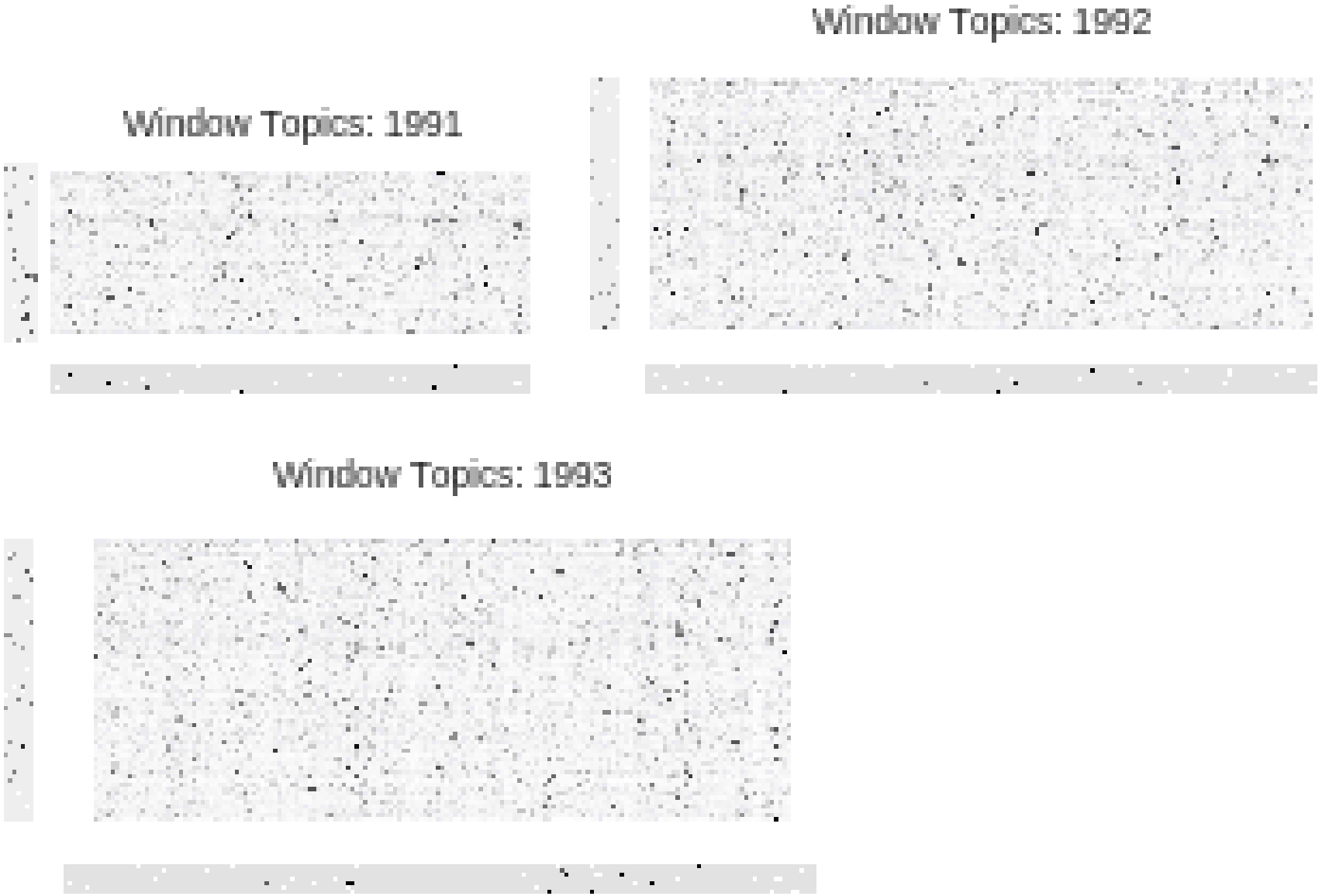
Where the size and variety of the literature we want to model has increased exponentially, we need an approach that allows for the emergence of new topics.

Dynamic NMF

Dynamic NMF



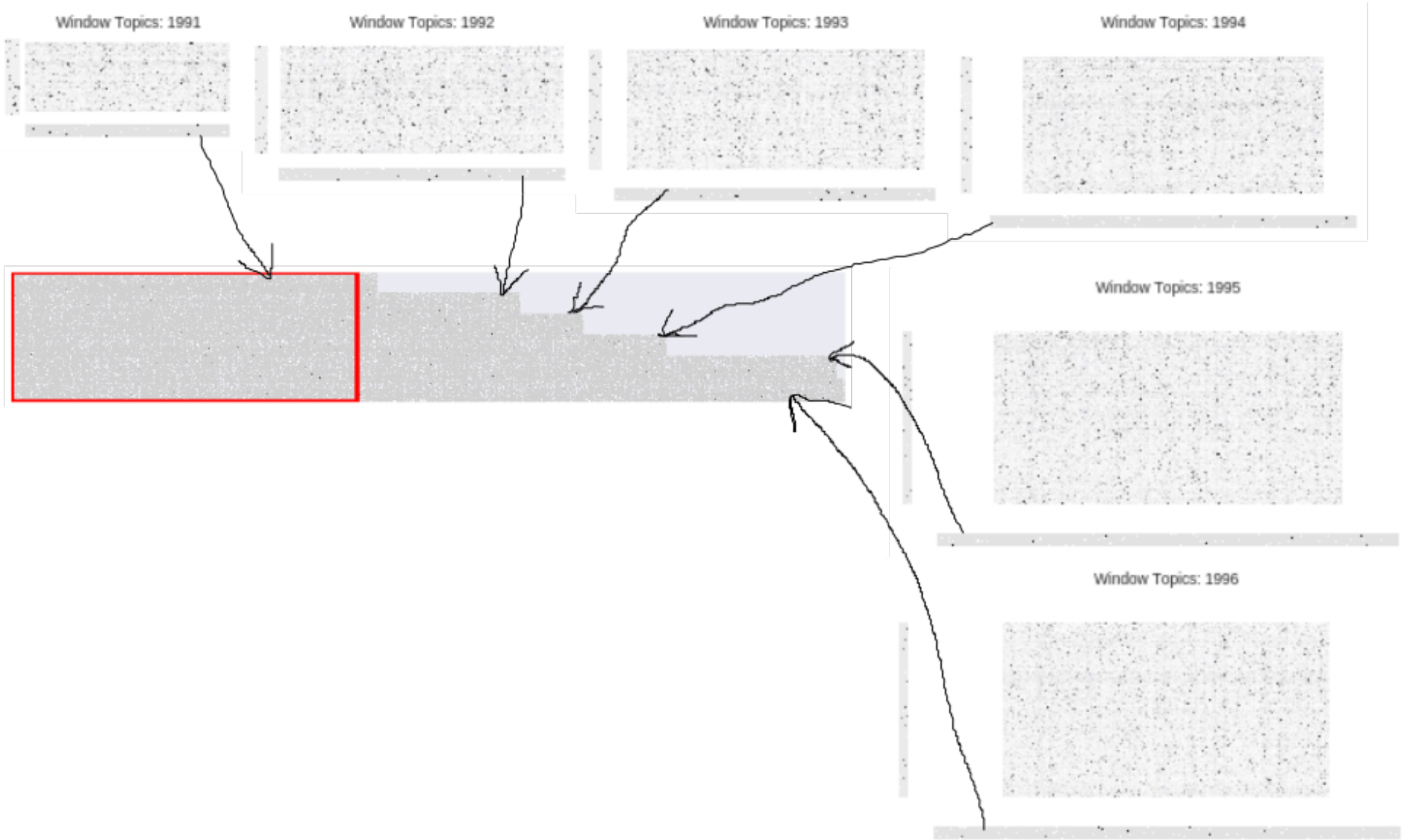
Dynamic NMF



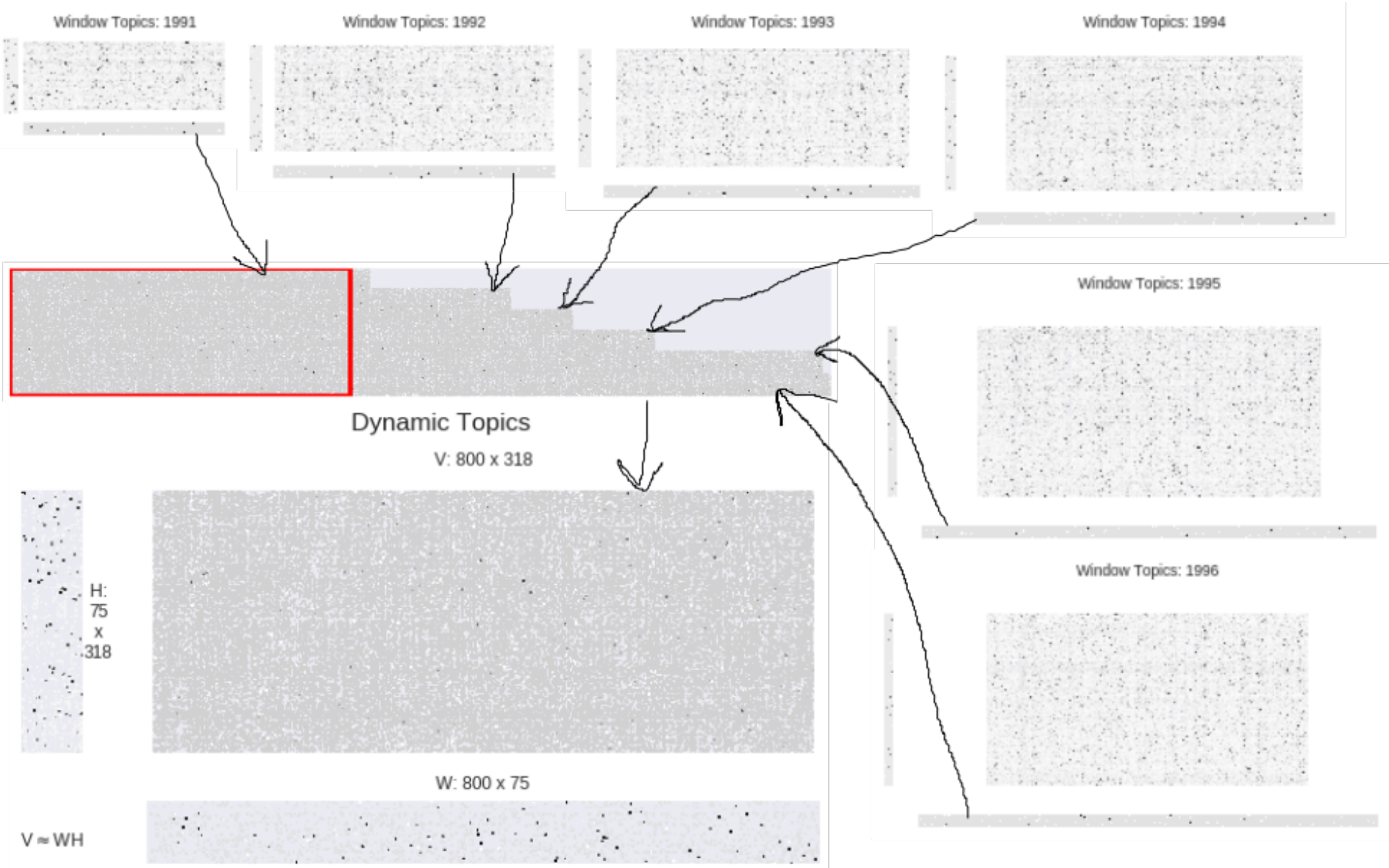
Dynamic NMF


Window Topics: 1991
Window Topics: 1992
Window Topics: 1993
Window Topics: 1994
Window Topics: 1995
Window Topics: 1996

Dynamic NMF


Window Topics: 1991
Window Topics: 1992
Window Topics: 1993
Window Topics: 1994
Window Topics: 1995
Window Topics: 1996

Dynamic NMF


Window Topics: 1991
Window Topics: 1992
Window Topics: 1993
Window Topics: 1994
Window Topics: 1995
Window Topics: 1996
Dynamic Topics
V: 800 x 318
H: 75 x 318
W: 800 x 75
V ≈ WH

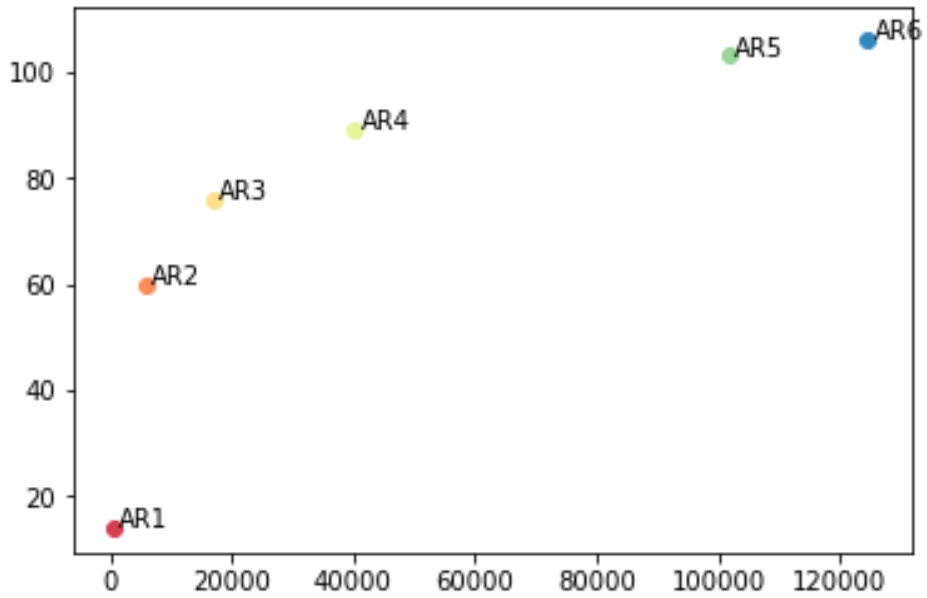
Dynamic NMF - application to climate change

- 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
- Predictive capacity

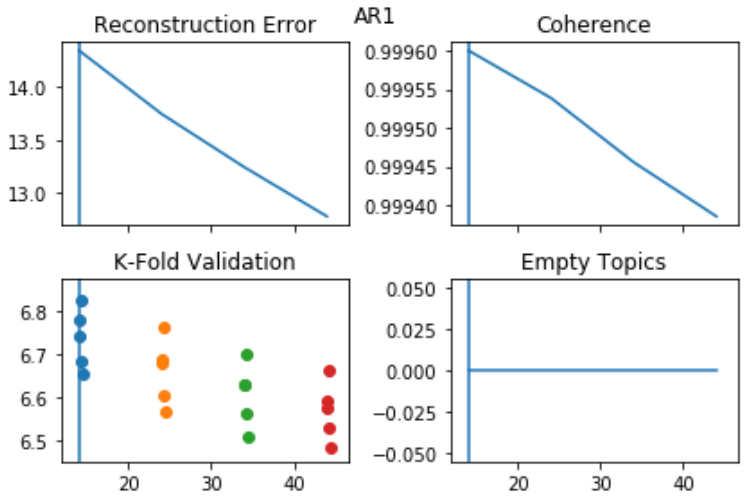
Dynamic NMF - number of topics

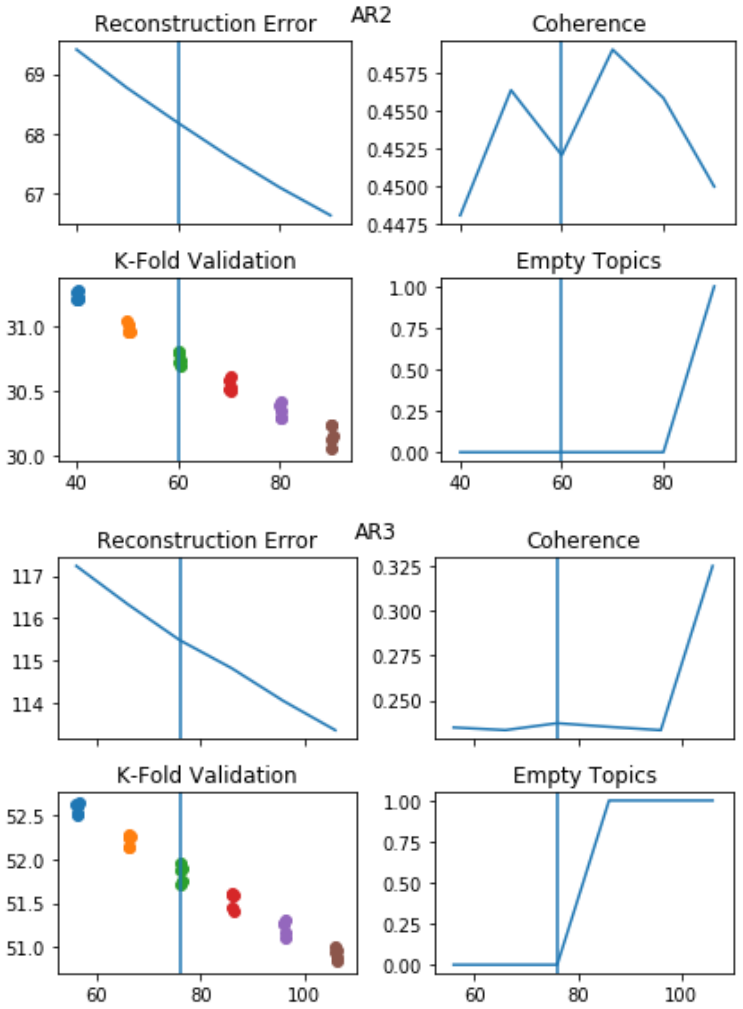
Human topic number criteria

- Intelligibility

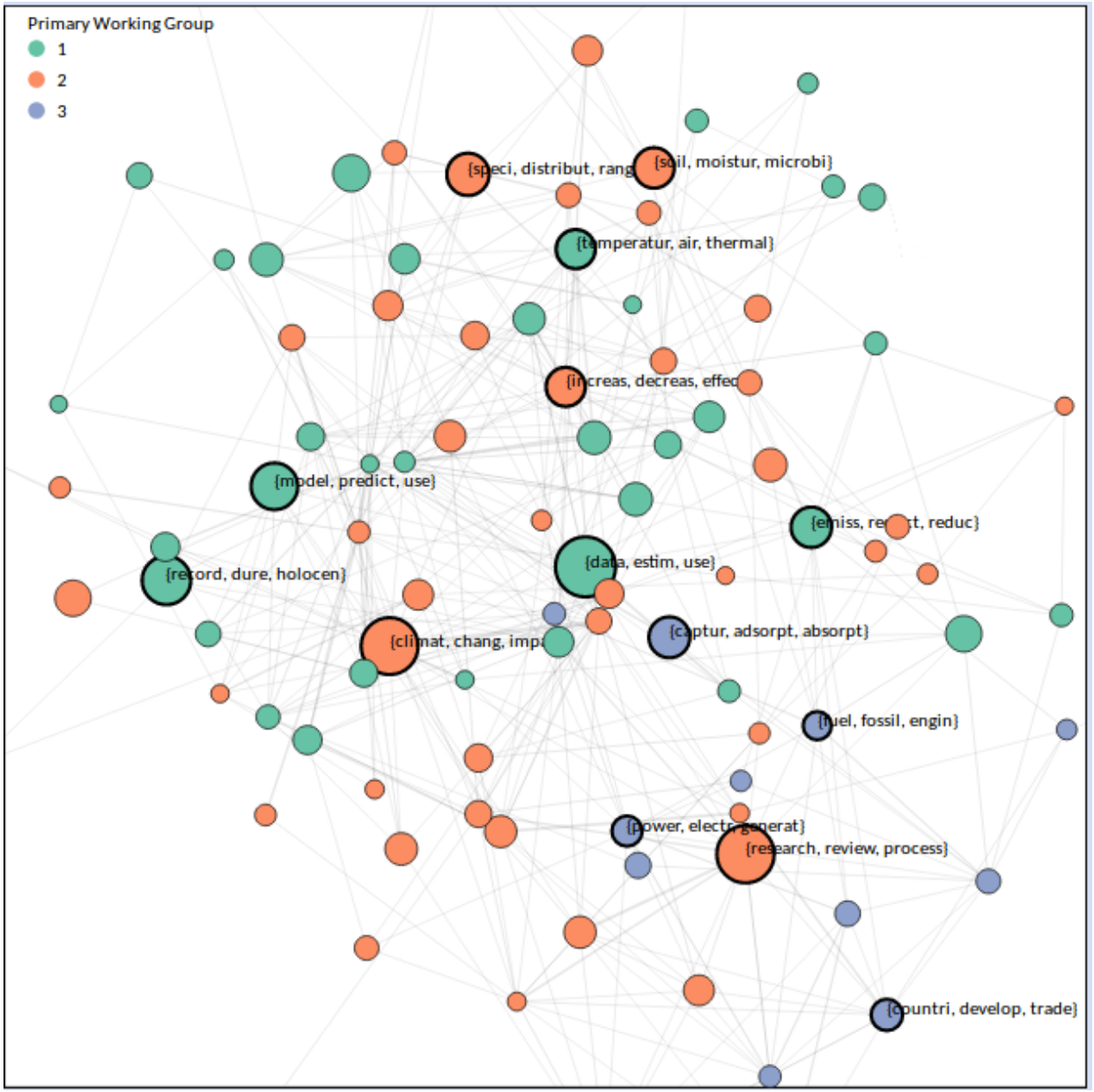
Data-driven topic number criteria

- Reconstruction accuracy
- Predictive capacity





Preliminary results - structure



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

IPCC Coverage	Primary WG	Topic Title	WG 1	WG 2	WG 3
0.16%	1	{rainfal, monsoon, rain}	0.50%	0.50%	0.00%
0.10%	2	{veget, ndvi, cover}	0.41%	0.59%	0.00%
0.16%	1	{snow, cover, winter}	0.59%	0.41%	0.00%
0.17%	2	{region, local, scale}	0.41%	0.59%	0.00%
0.16%	1	{coastal, mangrov, rise}	0.57%	0.42%	0.01%

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

IPCC Coverage	Primary WG	Topic Title	WG 1	WG 2	WG 3
0.09%	3	{gas, coal, greenhous}	0.30%	0.15%	0.56%
0.10%	3	{transport, vehicl, road}	0.24%	0.12%	0.64%
0.13%	1	{emiss, reduct, reduc}	0.45%	0.21%	0.34%
0.09%	1	{methan, oxid, methanotroph}	0.63%	0.16%	0.20%
0.13%	3	{ghg, greenhous, gas}	0.15%	0.09%	0.75%

Preliminary results - Inter-working Group topics - WG 2 and 3

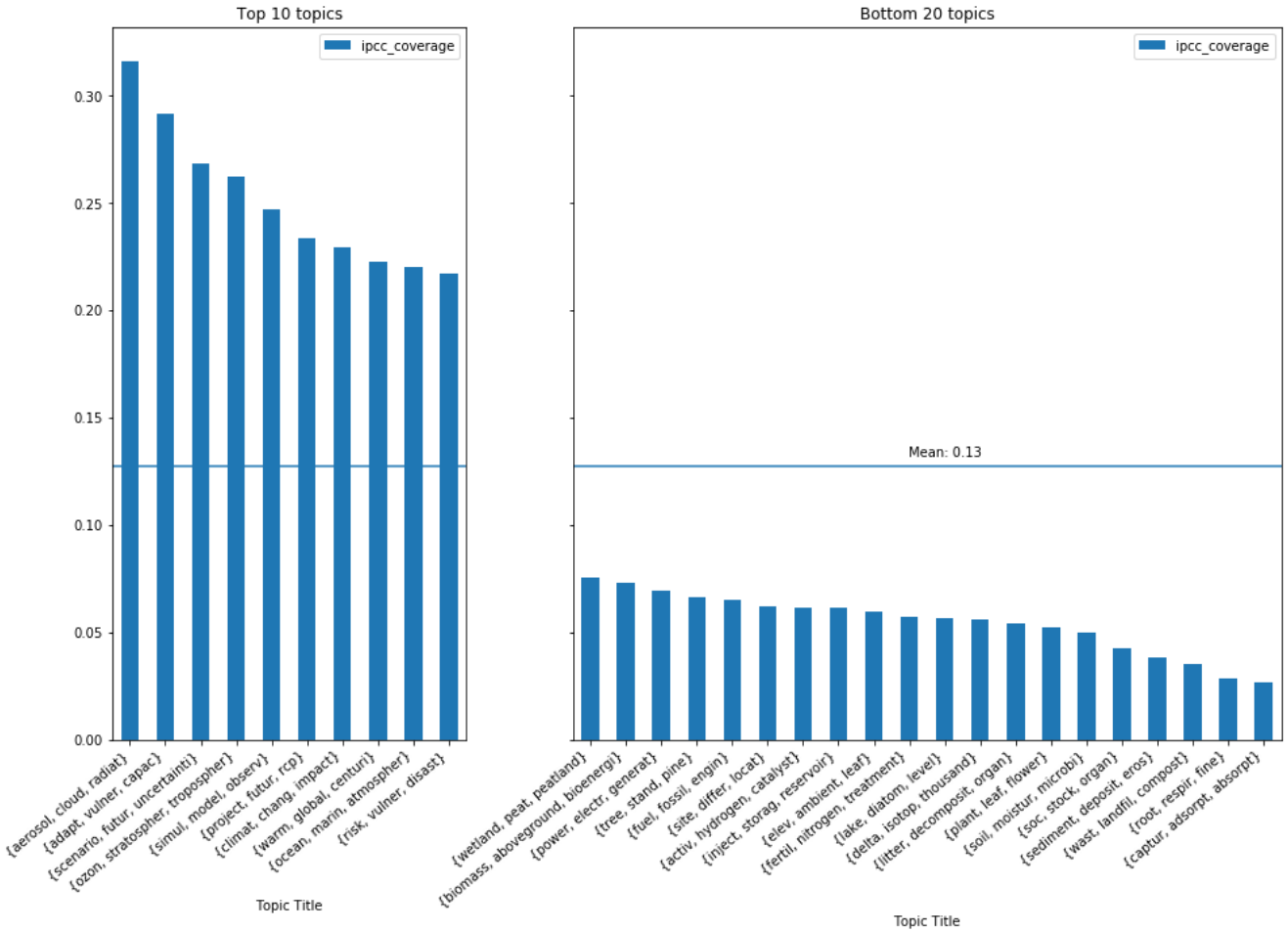
IPCC Coverage	Primary WG	Topic Title	WG 1	WG 2	WG 3
0.11%	2	{sustain, develop, resourc}	0.04%	0.51%	0.46%
0.08%	3	{build, construct, design}	0.03%	0.38%	0.59%
0.11%	2	{environment, impact, life}	0.06%	0.58%	0.36%
0.19%	3	{polici, tax, govern}	0.02%	0.32%	0.66%
0.16%	2	{urban, citi, plan}	0.07%	0.55%	0.38%

Preliminary results - gaps in coverage

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

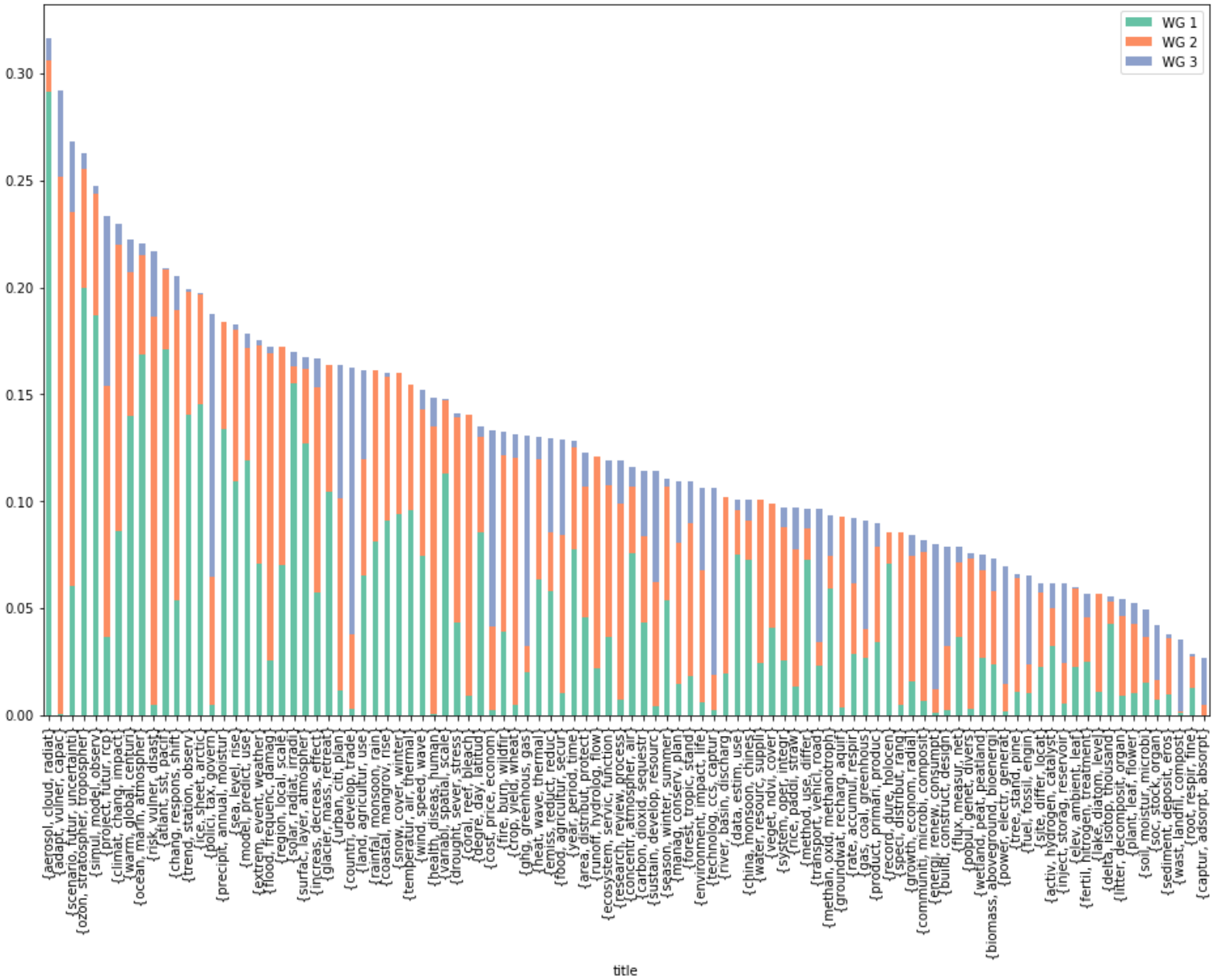
- Each document d either matches or does not match an IPCC reference
- For each topic h, we can sum the scores for each category of document
- The "IPCC proportion" of each topic is the proportion of the sum of the document score accounted for by documents which match IPCC references.

Preliminary results - gaps in coverage

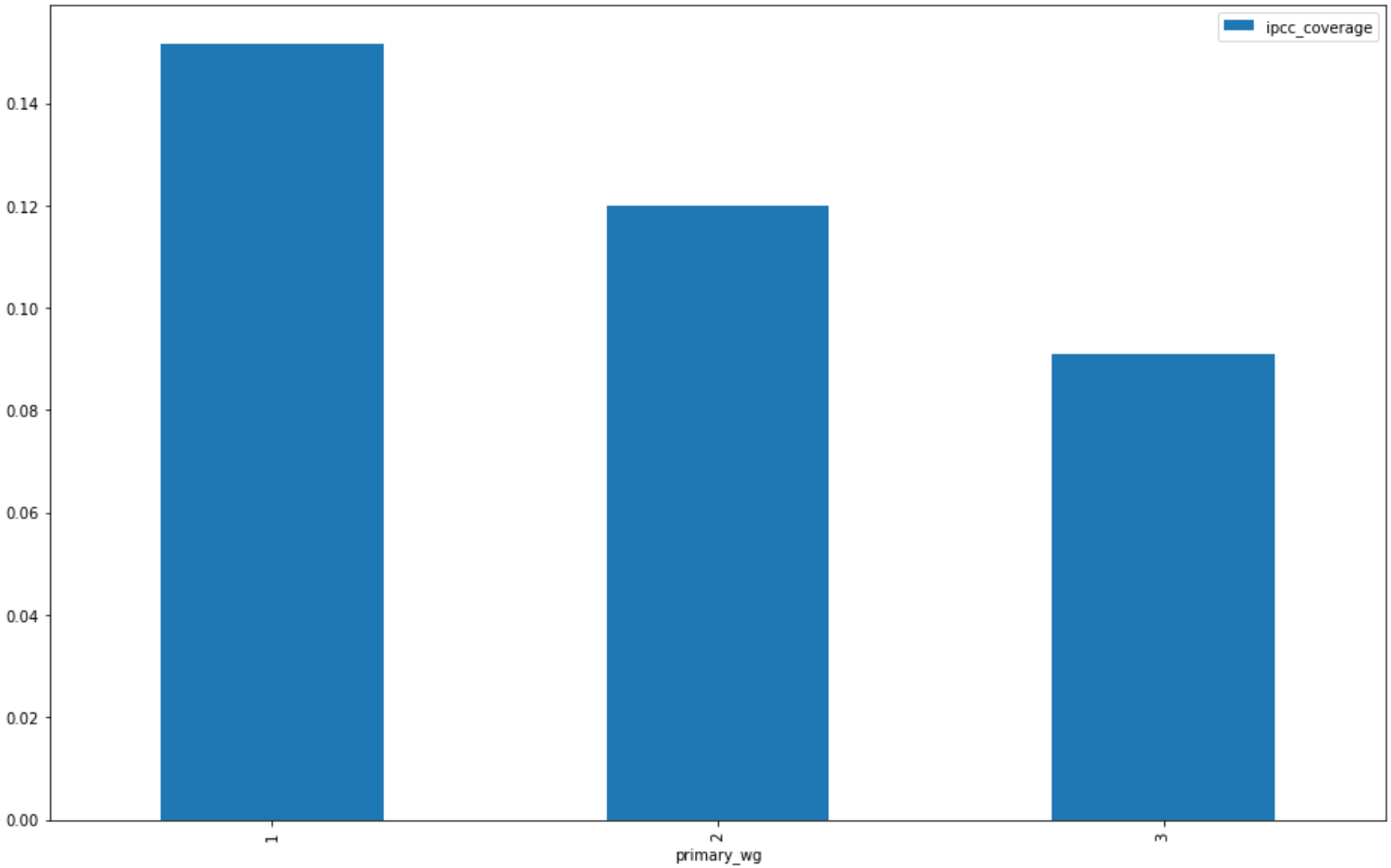


- The physical science aspects of climate change, as well topics on impacts, adaptation and scenarios are well covered by the IPCC
- Topics on specific technological solutions (particularly NETs), as well as soils, are less well covered

Preliminary results - gaps in coverage



Preliminary results - gaps in coverage



Next steps

- Adjusting second stage of model to give more weight to topics with many documents (to better capture emergence)
- Further analysis of new model
- (Manual) comparison of (AR6) topics with AR6 outline

Blei, D. M. and Lafferty, J. D. (2006). Dynamic Topic Models. *International Conference on Machine Learning*, page 113âĂŞ120.

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.

Haunschild, R., Bornmann, L., and Marx, W. (2016). Climate Change Research in View of Bibliometrics. *PLoS ONE*, 11(7):1–19.

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