A Topography of Climate Change Research

Max Callaghan

February 21, 2018

Introduction

Figure: Portrait of map-makers, Gerard Mercator and Jodocus Hondius (Jodocus Hondius) source: `https://commons.wikimedia.org/wiki/File:Hondius_Portrait_of_map-makers.jpg`

Introduction

Figure: Portrait of map-makers, Gerard Mercator and Jodocus Hondius (Jodocus Hondius) source: `https://commons.wikimedia.org/wiki/File:Hondius_Portrait_of_map-makers.jpg`

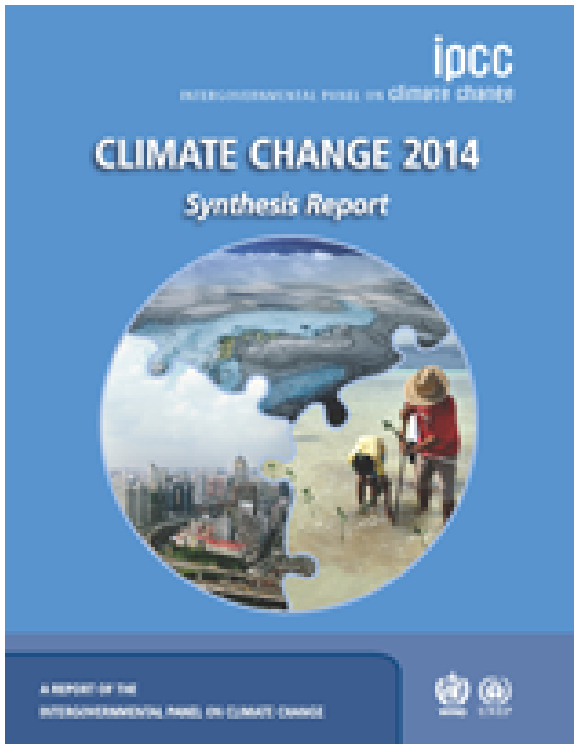
- Topography is a description of a landscape

Introduction

Figure: Portrait of map-makers, Gerard Mercator and Jodocus Hondius (Jodocus Hondius) source: `https://commons.wikimedia.org/wiki/File:Hondius_Portrait_of_map-makers.jpg`

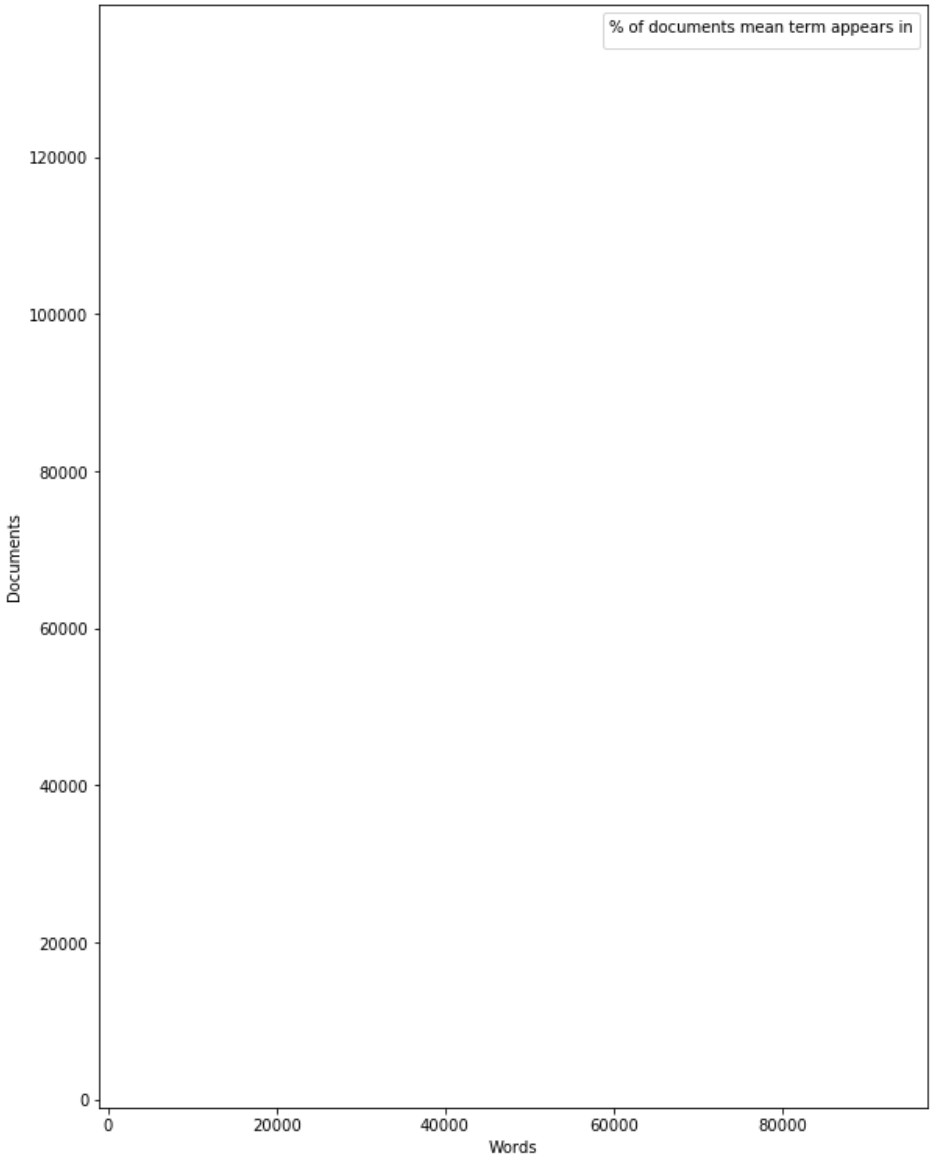
- Topography is a description of a landscape
- Topics (from the Greek τόπος, place) can describe the features of body of text

Context



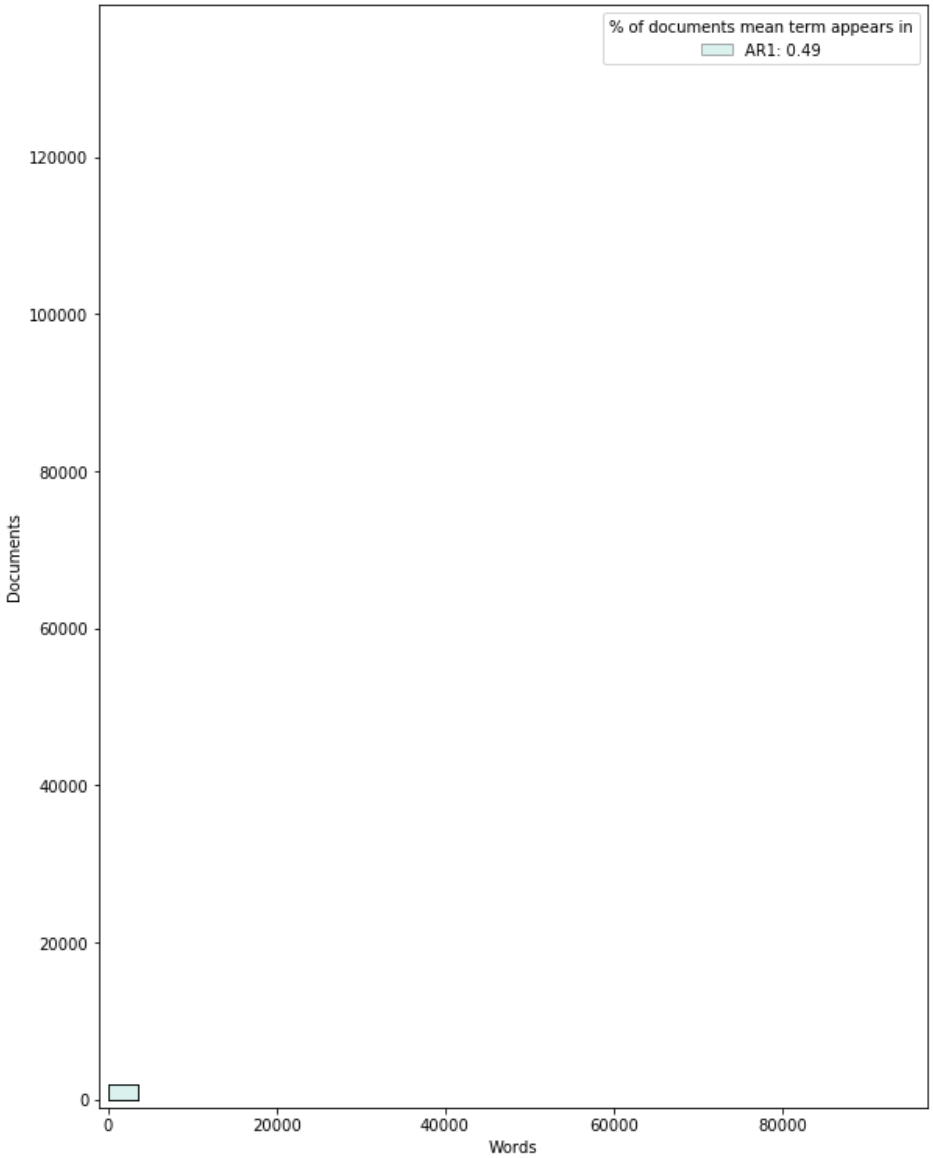
- To contribute evidence-based policy-making on climate change, the IPCC aims to *comprehensively* assess scientific literature on climate change
- These assessments should be aim to balance legitimacy, credibility and relevance (Cash and Clark, 2001)

Motivation



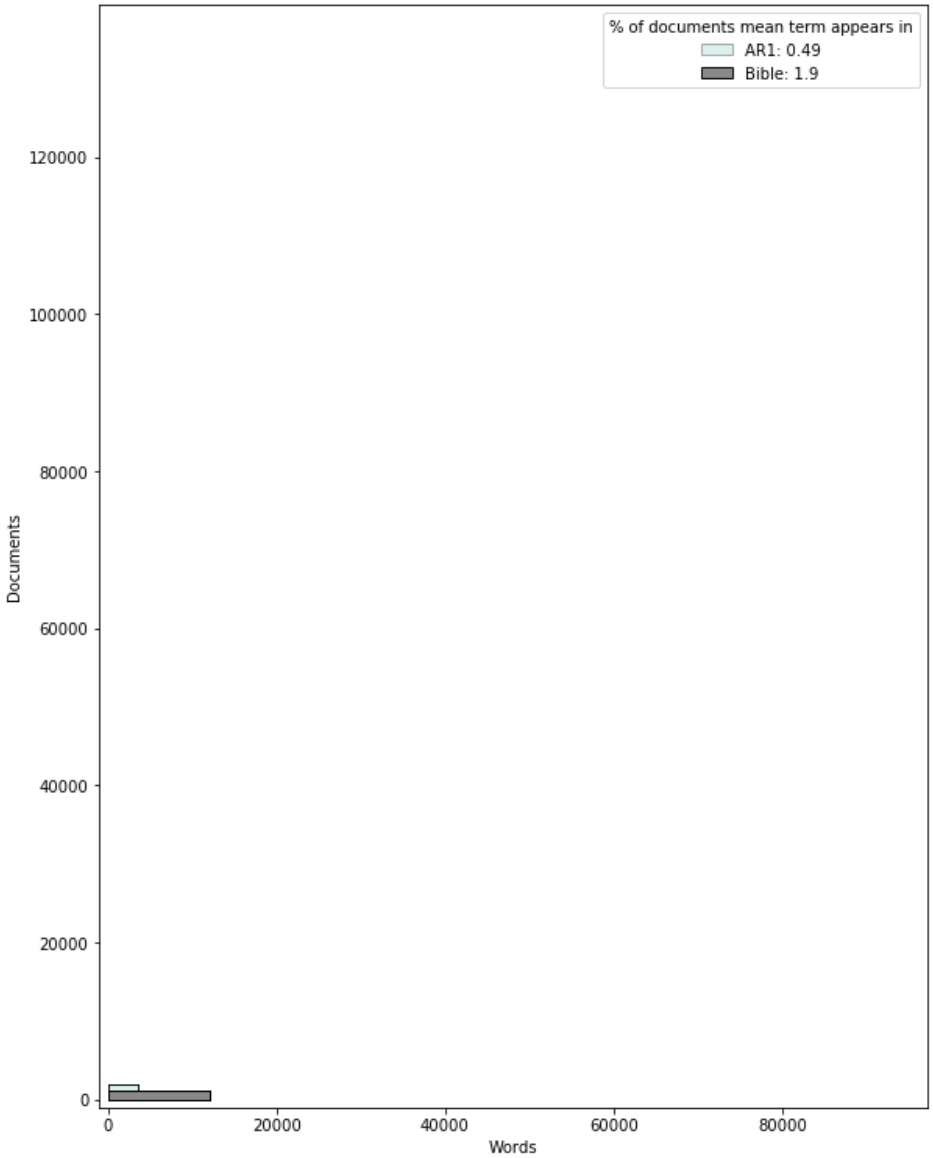
A matrix of documents x words

Motivation



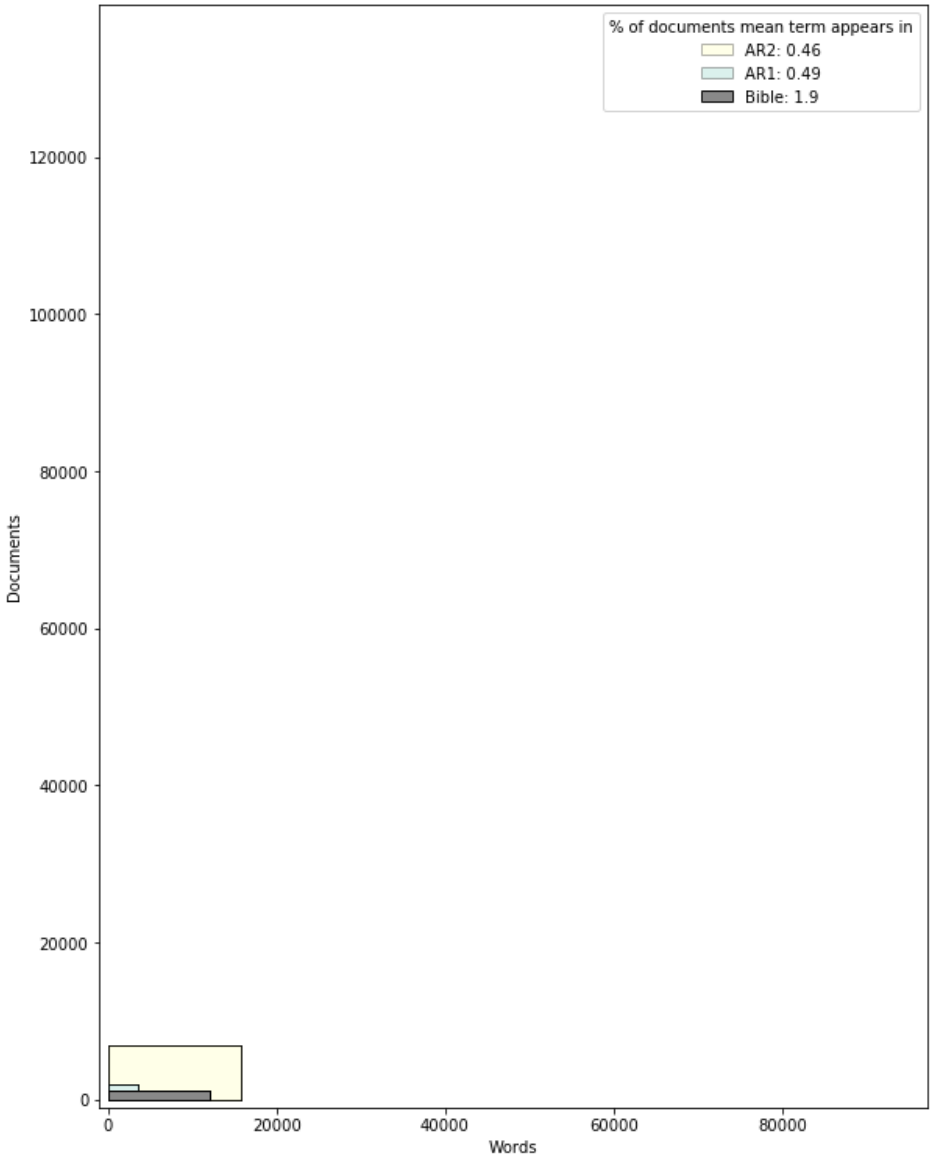
AR1: 1,848 documents x 3,528 words

Motivation



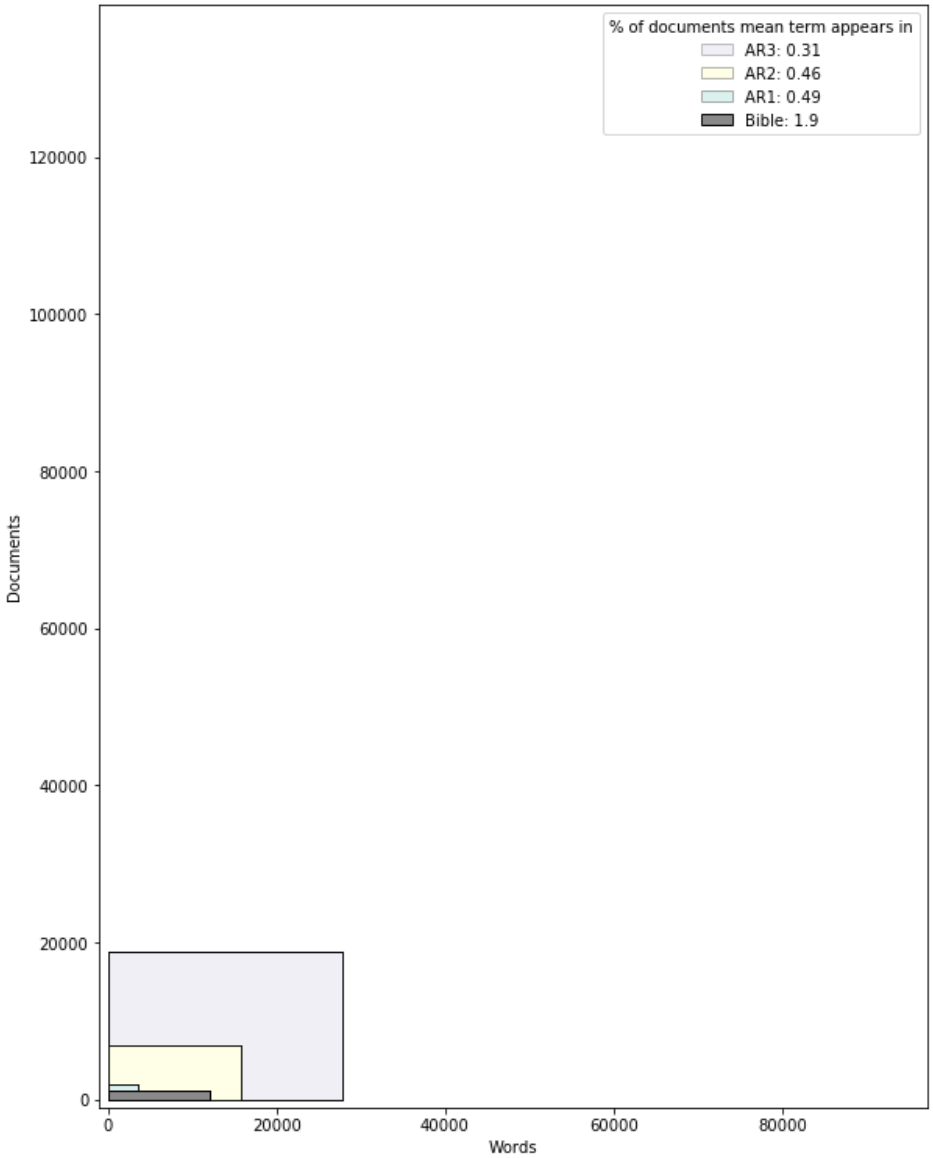
The Luther Bible: 1,189 documents (chapters) x 11,973 words

Motivation



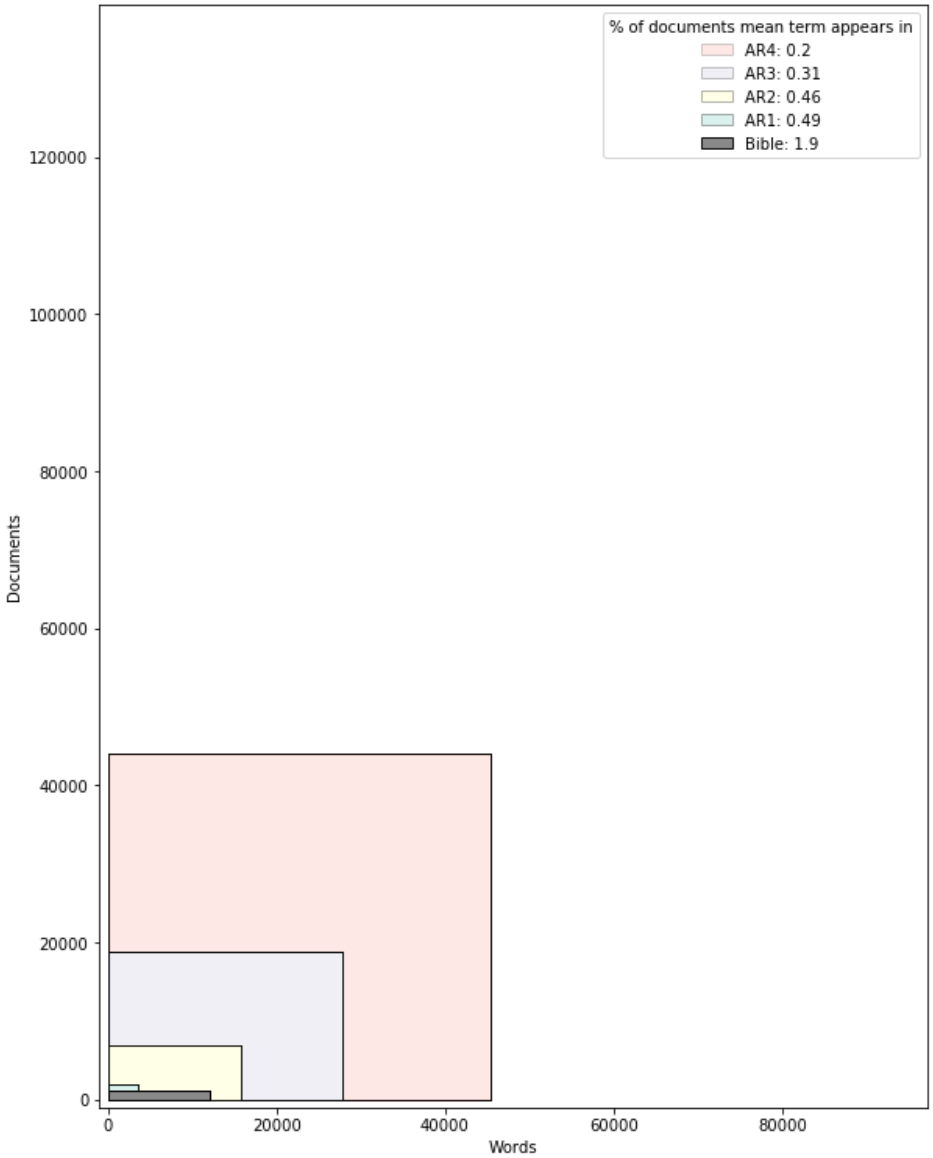
AR2: 6,941 documents x 15,781 words

Motivation



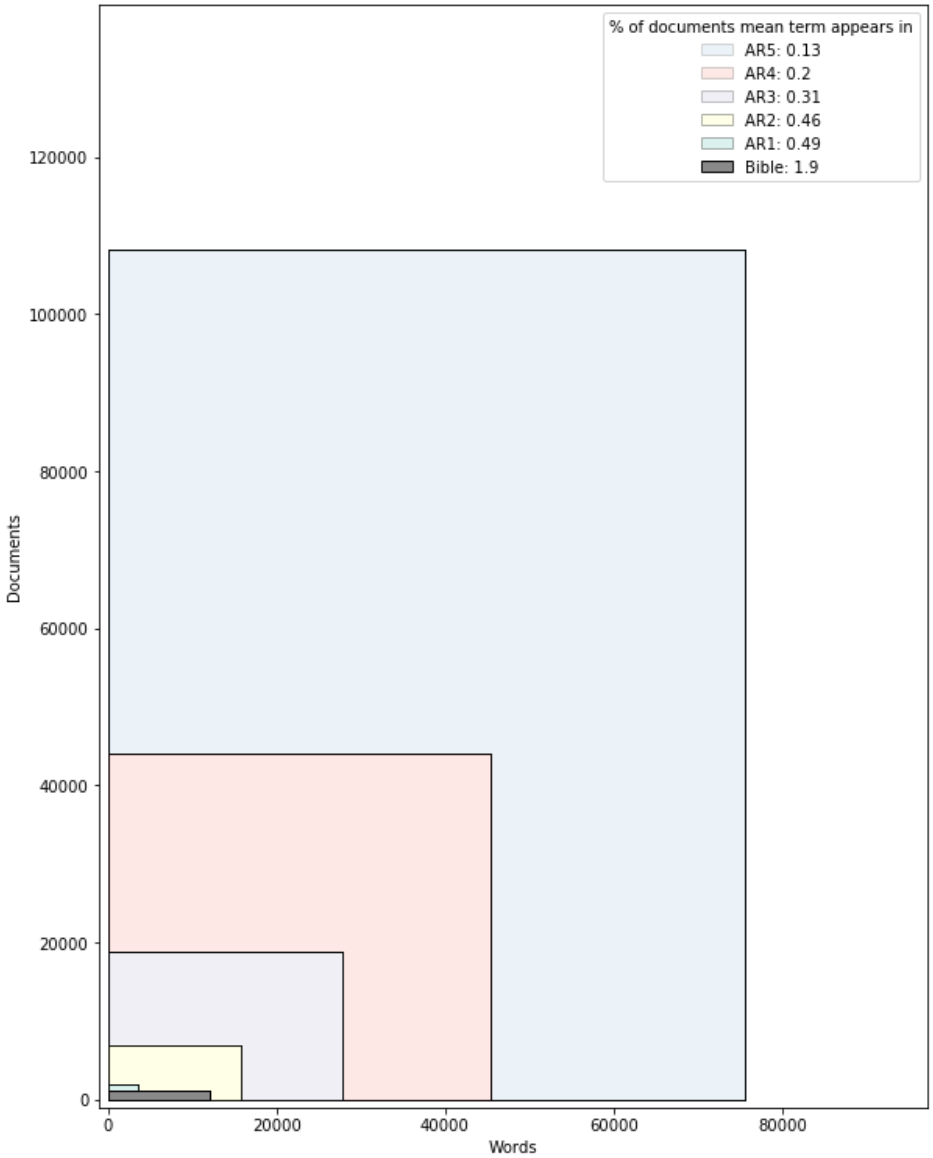
AR3: 18,728 documents x 27,730 words

Motivation



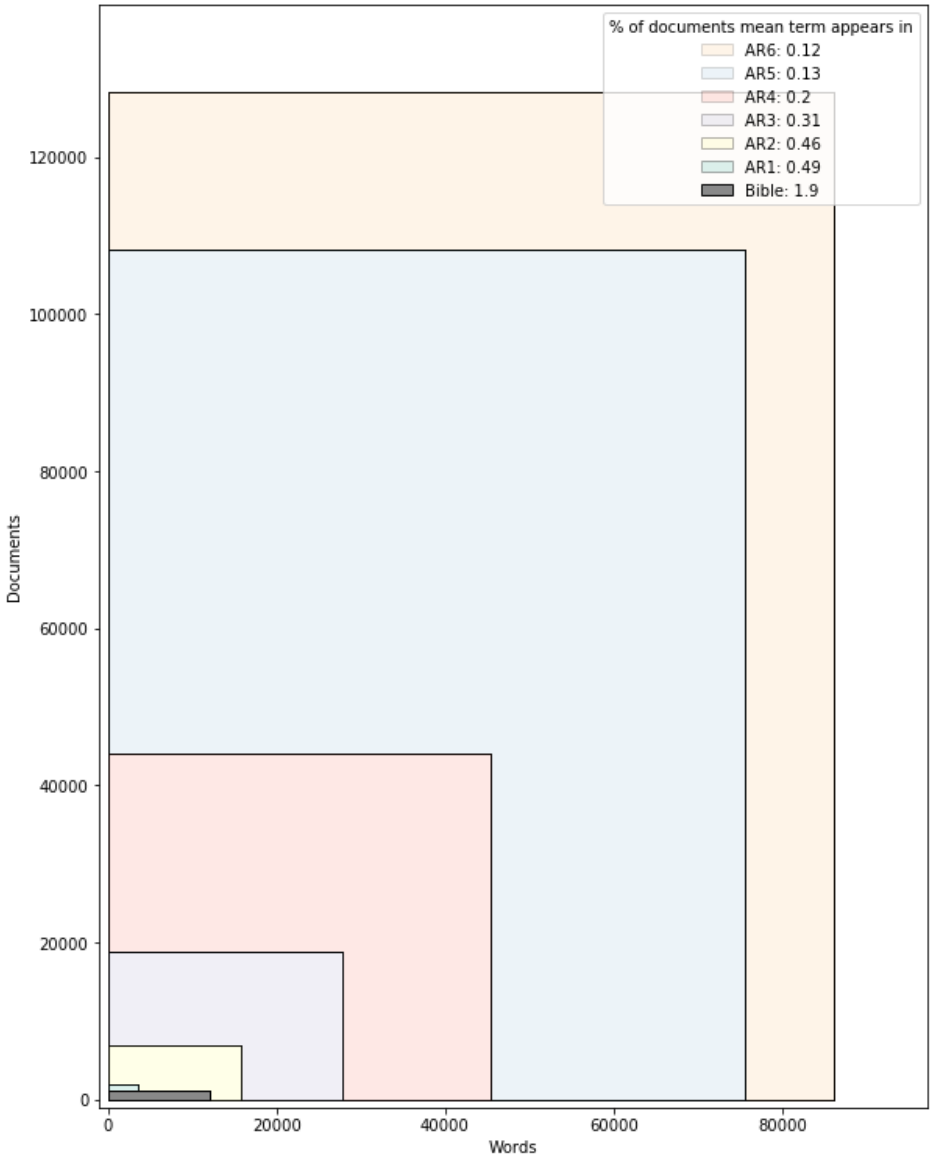
AR4: 44,000 documents x 45,388 words

Motivation



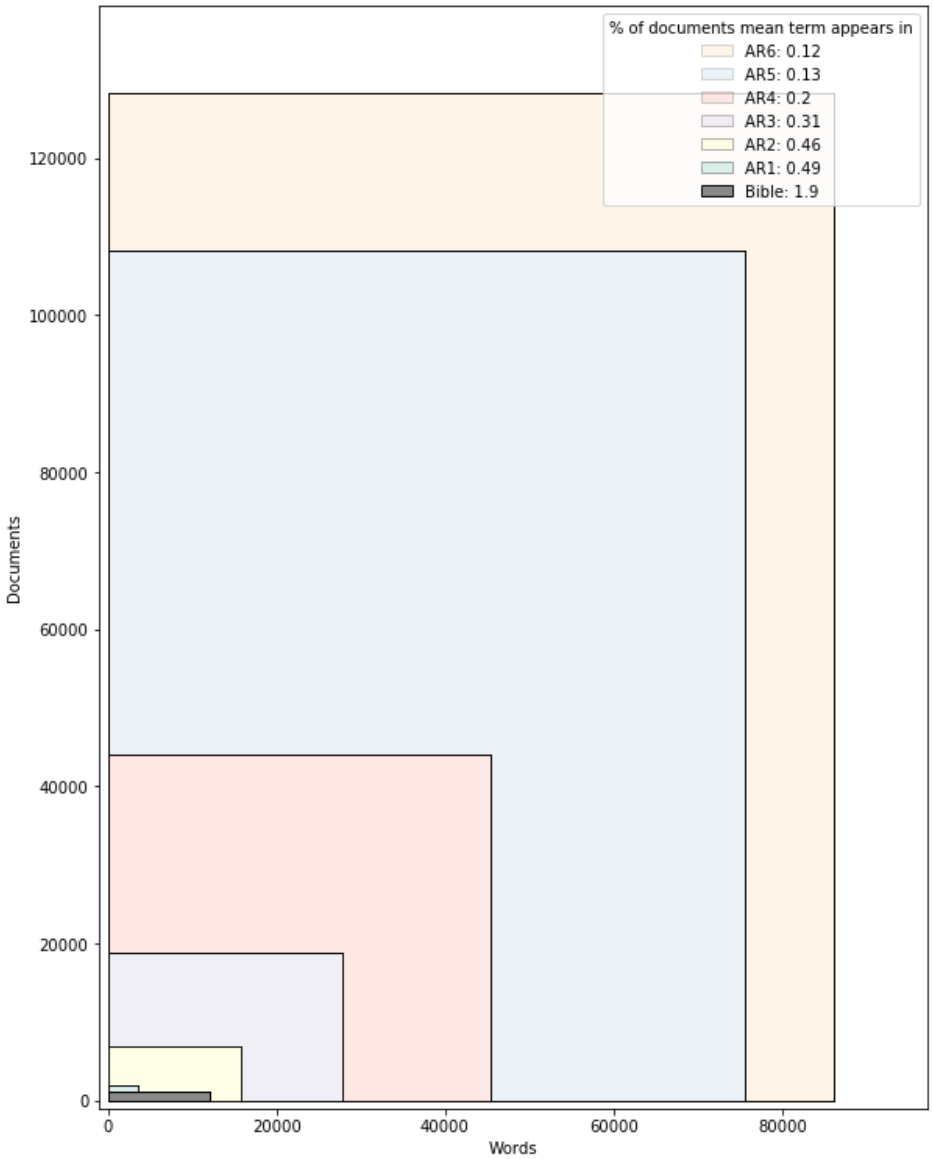
AR5: 108,277 documents x 75,553 words

Motivation



AR6: 128,357 documents x 86,149 words

Motivation

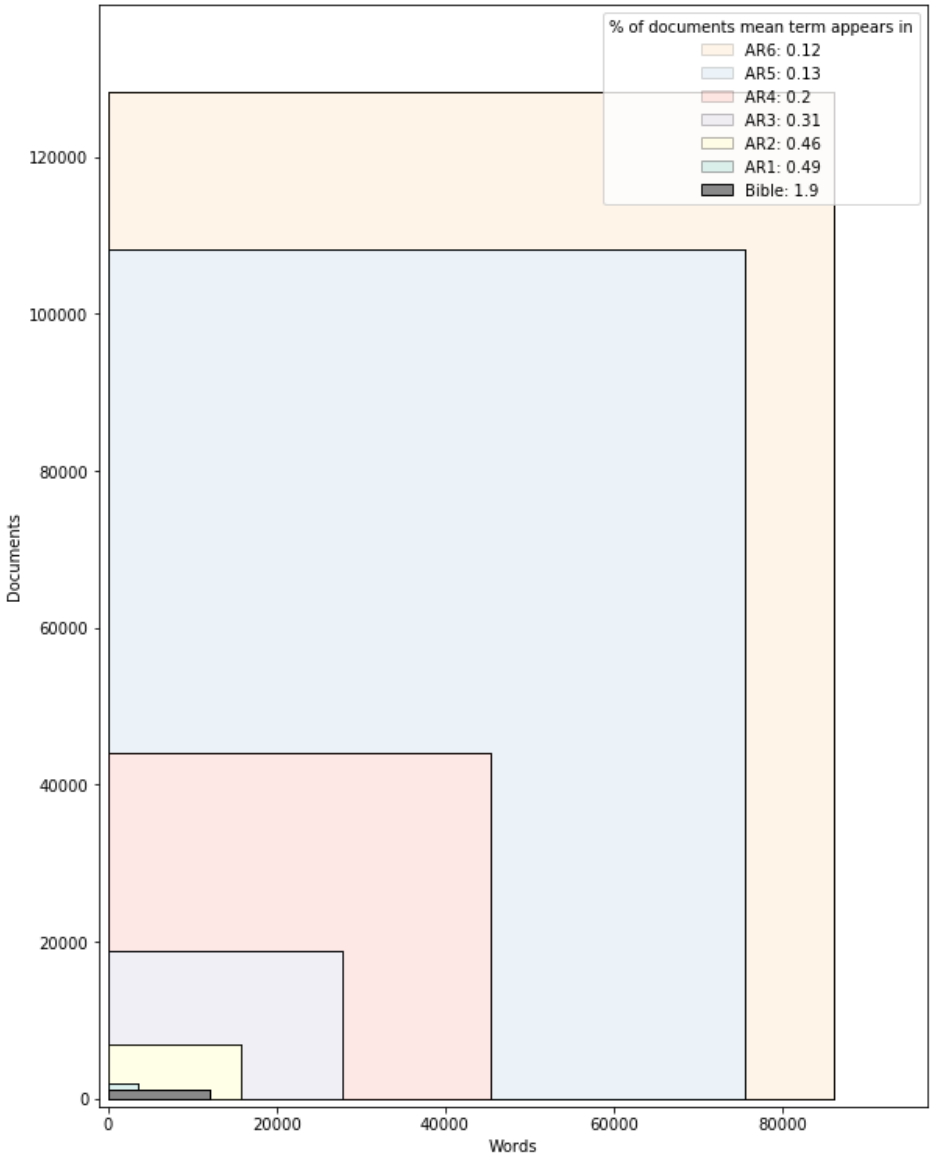


AR6: 128,357 documents x 86,149 words

- 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 (What is the matter?)

Approach - Words, words, words

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

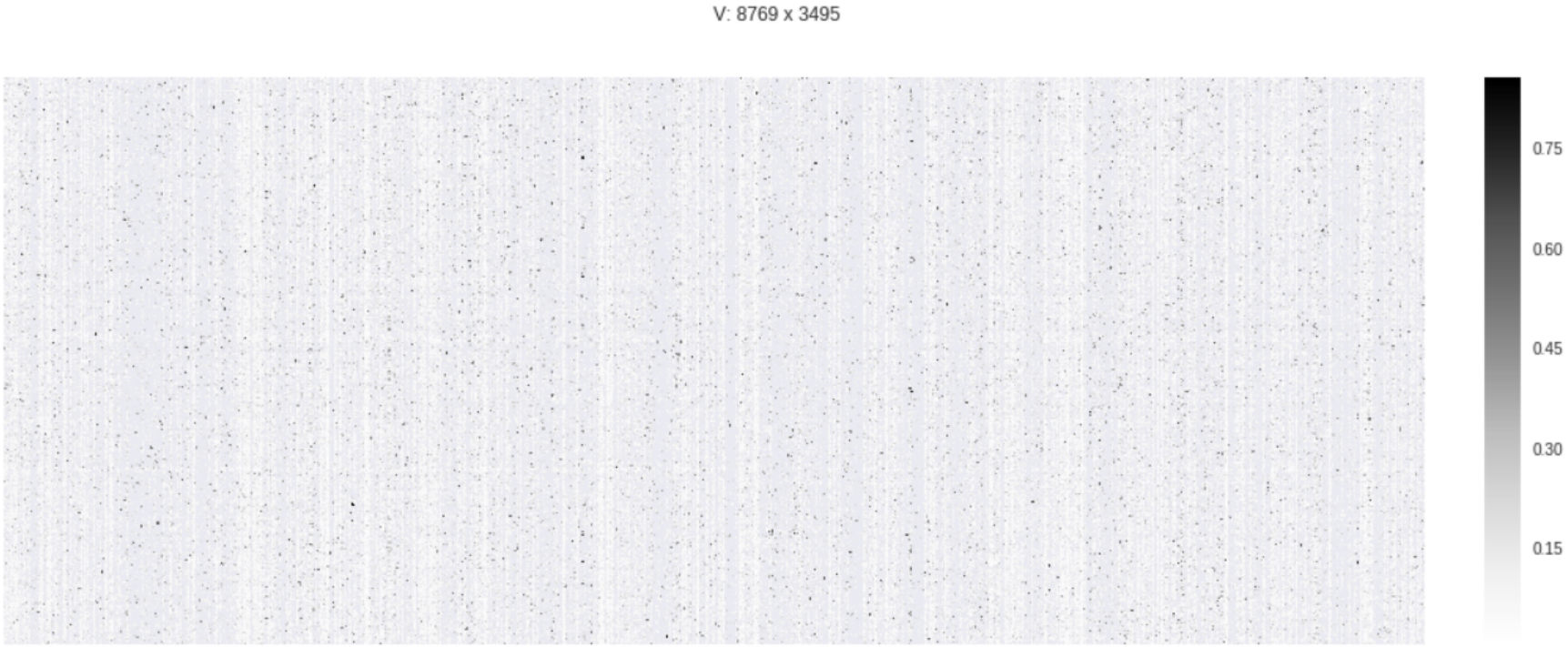


Figure: A topic model of 3495 documents on climate change from the year 2000

Approach - Words, words, words

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

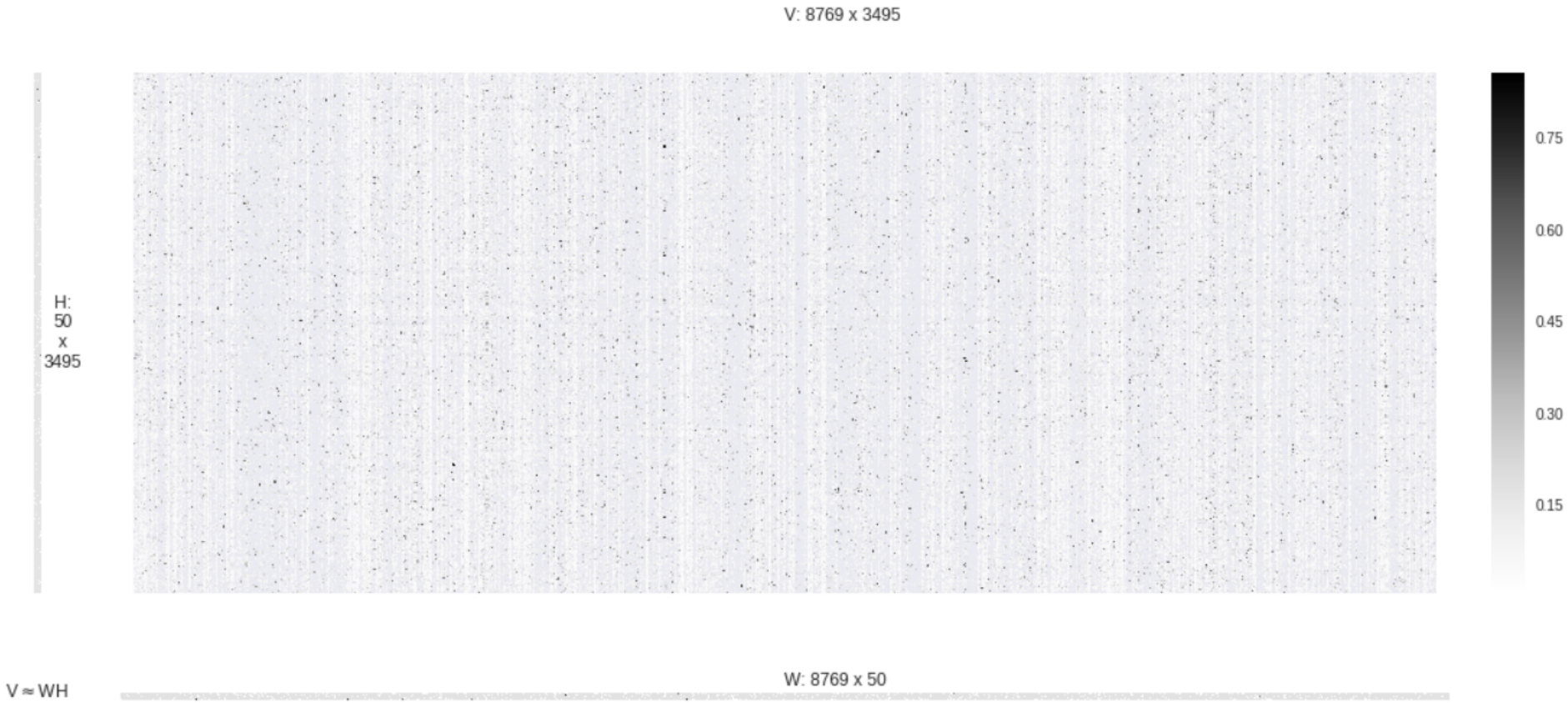


Figure: A topic model of 3495 documents on climate change from the year 2000

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

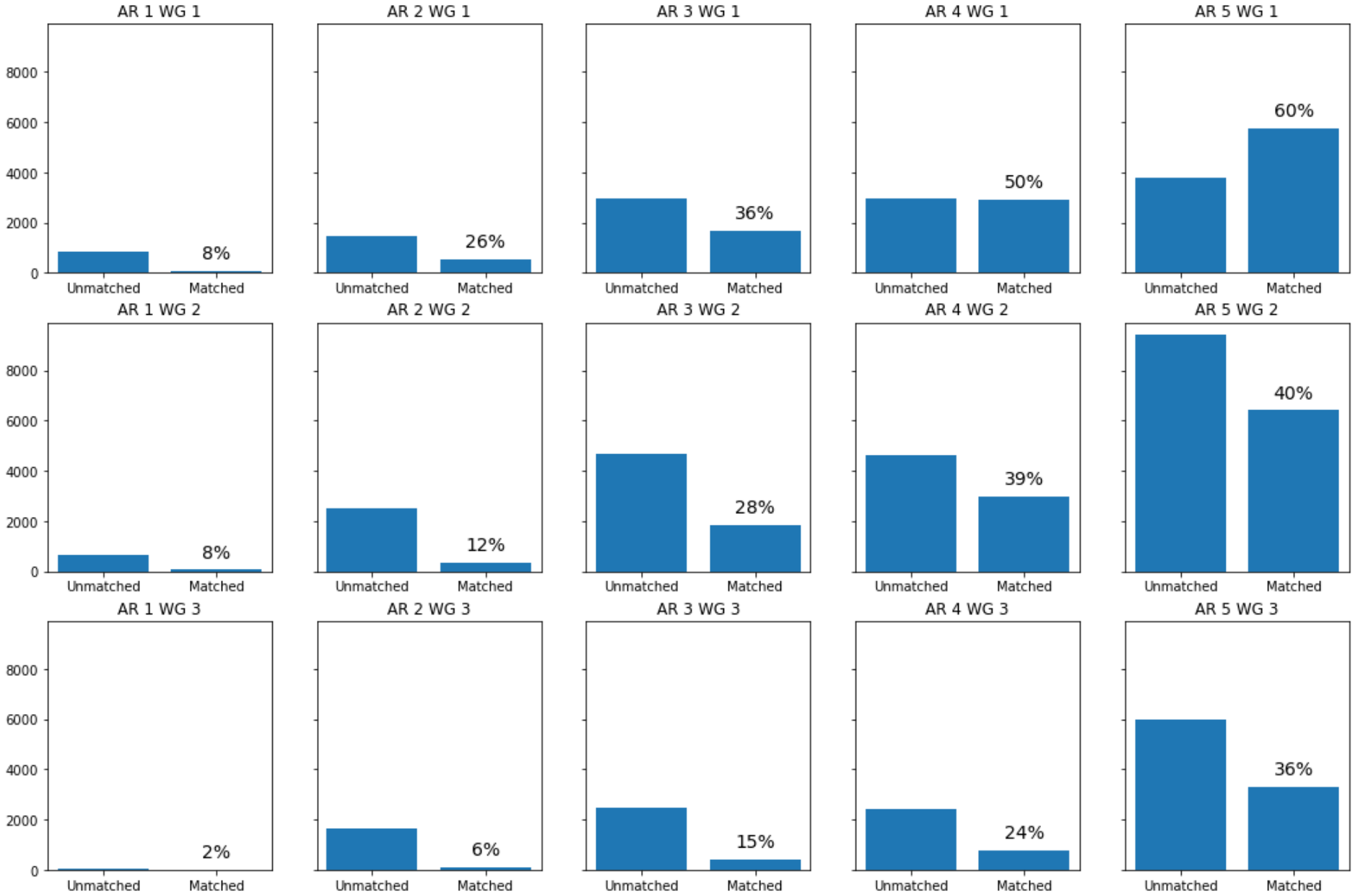
MCC

Matching process

For each Reference:

- Check for case-insensitive title matches
- Calculate the Jaccard similarity score for two word every database document containing the first word and from the same year. Match if the Jaccard score is above 0.45

Data - IPCC References



Data - Unmatched IPCC References

AR	WG	text	authors	year
2	2	Landfill gas: working with Gaia. Biodeterioration Extracts no. 4, Energy Technology Support Unit, Harwell Laboratory, Oxfordshire, UK.	Richards, K.M.	1989
1	1	Longwave cloud radiative forcing as determined from Nimbus-7 observations J Cltm , 2, 766 799	Ardanuy, P E , L L Stowe, A Gruber, M Weiss and C S Long	1989
3	2	Climate change: overview and implications for wildlife. In: Wildlife Responses to Climate Change: North American Case Studies [Schneider, S.H. and T.L. Root (eds.)]. Island Press, Washington, DC, USA, (in press).	Root, T.L. and S.H. Schneider	2001
2	3	The impact ol' global wanning on the United States: A survey of recent literature, mimco. Institute lor Interna-tional Economics, Washington, DC (April).	Cline, W.R.	1993
3	2	Population-environment relations at the forested frontier of Nepal. Applied Geography, 20, 221-242.	Conway, D., K. Bhattarai, and N.R. Shrestha	2000
3	2	The Cities Project. Australian Geological Survey Organisation, Australia. Available online at http://www.agso.gov.au/geohazards/grm/cities2.html.	AGSO	1999
4	1	CLIMBER-2: A climate system model of intermediate complexity. Part I: Model description and performance for present climate.	Petoukhov, V., et al.	2000
5	1	A skill-score based evaluation of simulated Australian climate. Australian Meteorol. Oceanogr	Watterson, I., A. C. Hirst, and L. D. Rotstayn	2013
5	1	Enhanced aerosol backscatter adjacent to tropical trade wind clouds revealed by satellite-based lidar. Geophys. Res	Tackett, J. L., and L. Di Girolamo	2009
5	3	Promoting long-term invest-ments by institutional investors. OECD Journal: Financial Market Trends 1, 145 – 164	Della Croce R, F Stewart, and J Yermo	2011

Data - Unmatched IPCC References

37% of IPCC References could be matched to the database of climate-relevant documents

Reasons for not matching

Data - Unmatched IPCC References

37% of IPCC References could be matched to the database of climate-relevant documents

Reasons for not matching

1. Reference is not in WoS (not peer-reviewed, in minor journal)

Data - Unmatched IPCC References

37% of IPCC References could be matched to the database of climate-relevant documents

Reasons for not matching

1. Reference is not in WoS (not peer-reviewed, in minor journal)
2. Reference is not directly about climate change

Data - Unmatched IPCC References

37% of IPCC References could be matched to the database of climate-relevant documents

Reasons for not matching

1. Reference is not in WoS (not peer-reviewed, in minor journal)
2. Reference is not directly about climate change
3. Reference has been incorrectly parsed

Data - Unmatched IPCC References

37% of IPCC References could be matched to the database of climate-relevant documents

Reasons for not matching

1. Reference is not in WoS (not peer-reviewed, in minor journal)
2. Reference is not directly about climate change
3. Reference has been incorrectly parsed

Observations

- The size of the literature appears to be *much* bigger than our estimate

Data - Unmatched IPCC References

37% of IPCC References could be matched to the database of climate-relevant documents

Reasons for not matching

1. Reference is not in WoS (not peer-reviewed, in minor journal)
2. Reference is not directly about climate change
3. Reference has been incorrectly parsed

Observations

- The size of the literature appears to be *much* bigger than our estimate
- WG3 refers to more literature not directly about climate change, or not in peer-reviewed publications, than WG2, which refers to more than WG1

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

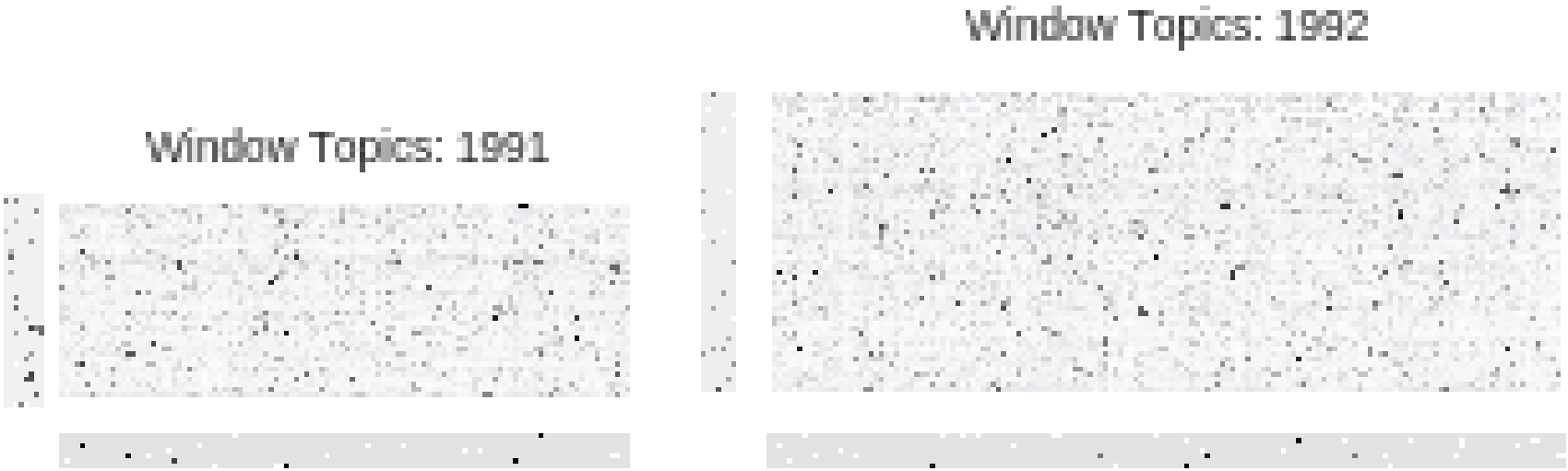
The topic models above assume that the topics, and the words that make them up, are stable over time. Two approaches to better model dynamic topics:

- 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.

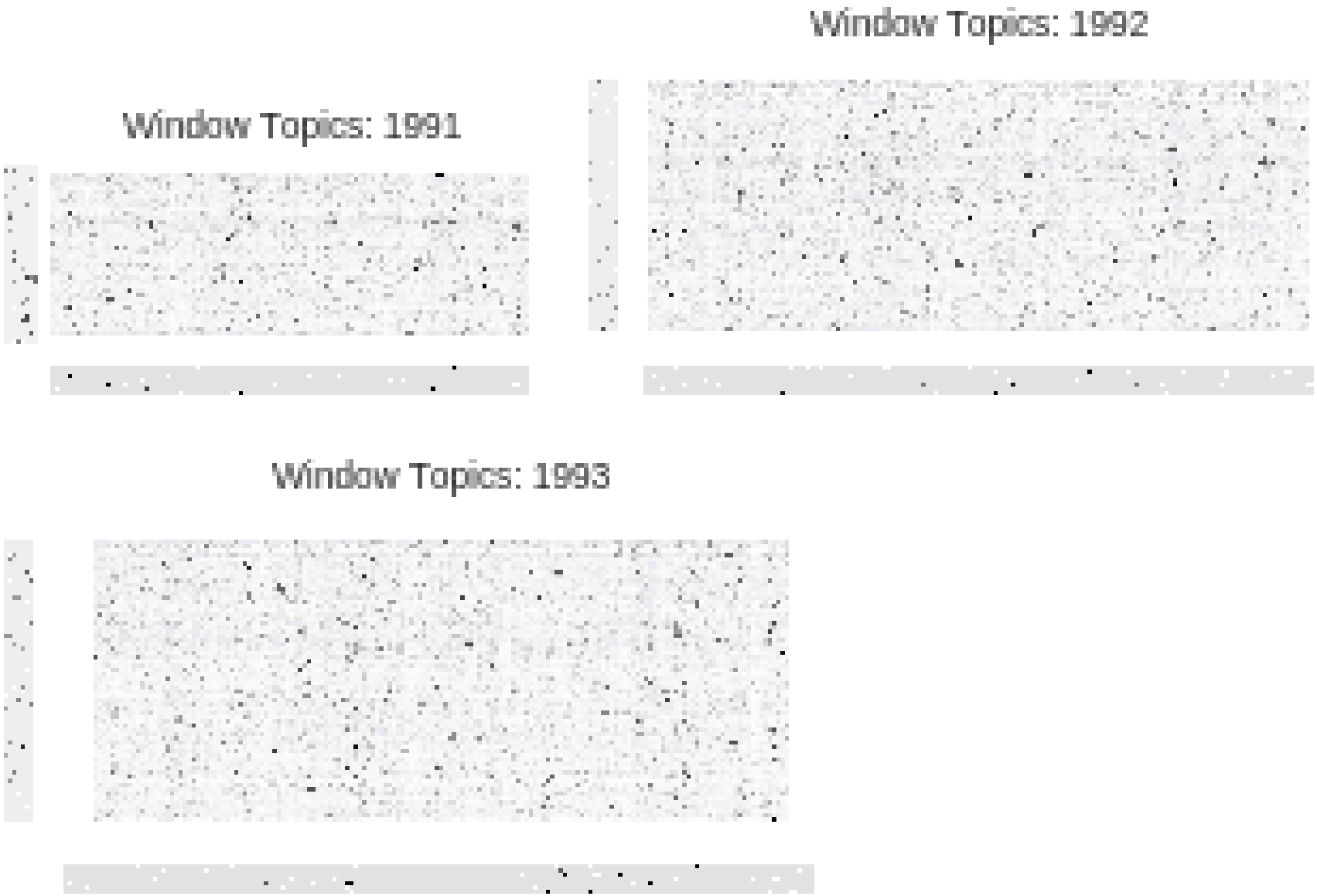
Where the size and variety of the literature we want to model has increased exponentially, we need an approach that allows for the emergence of new topics.

Dynamic NMF

Dynamic NMF



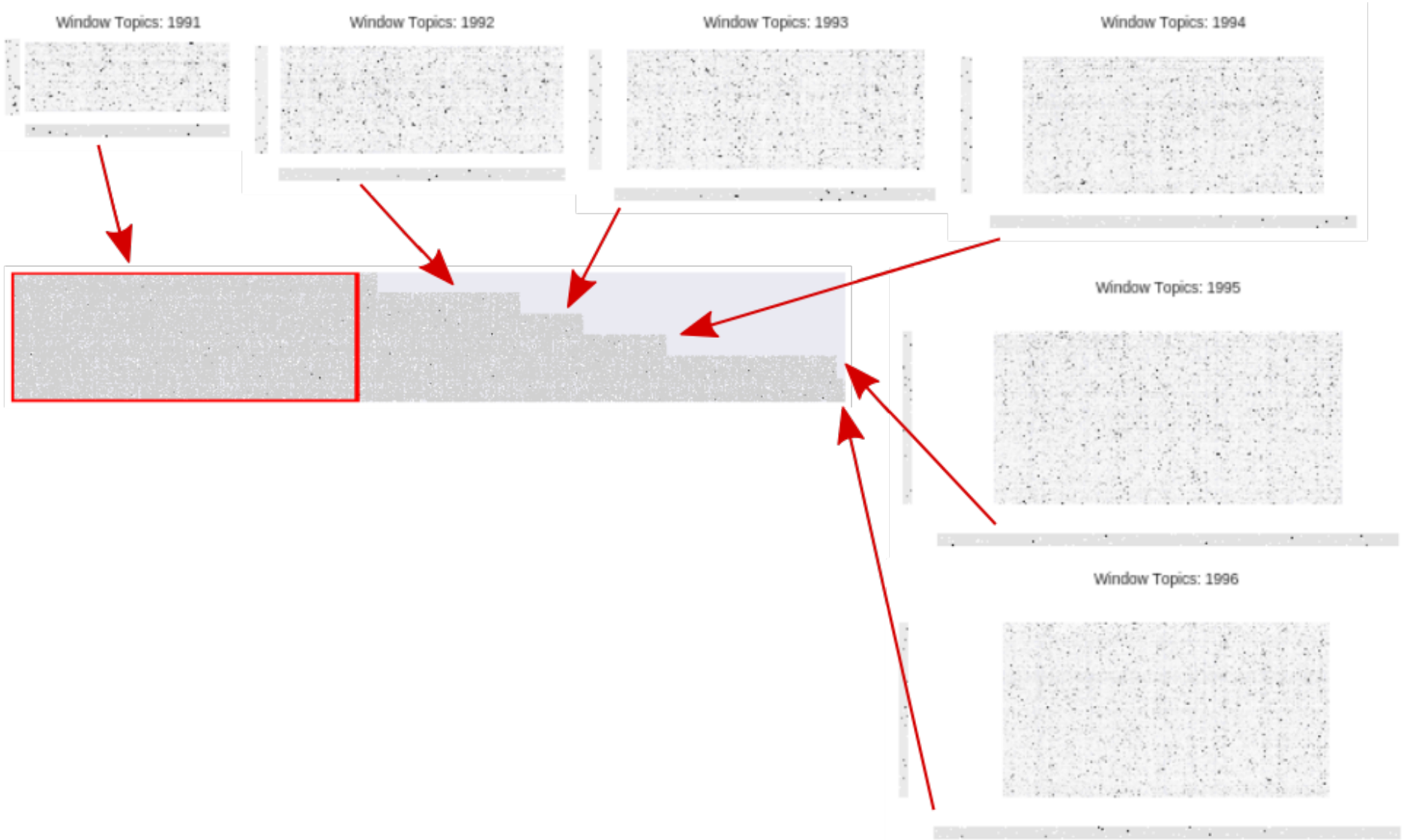
Dynamic NMF



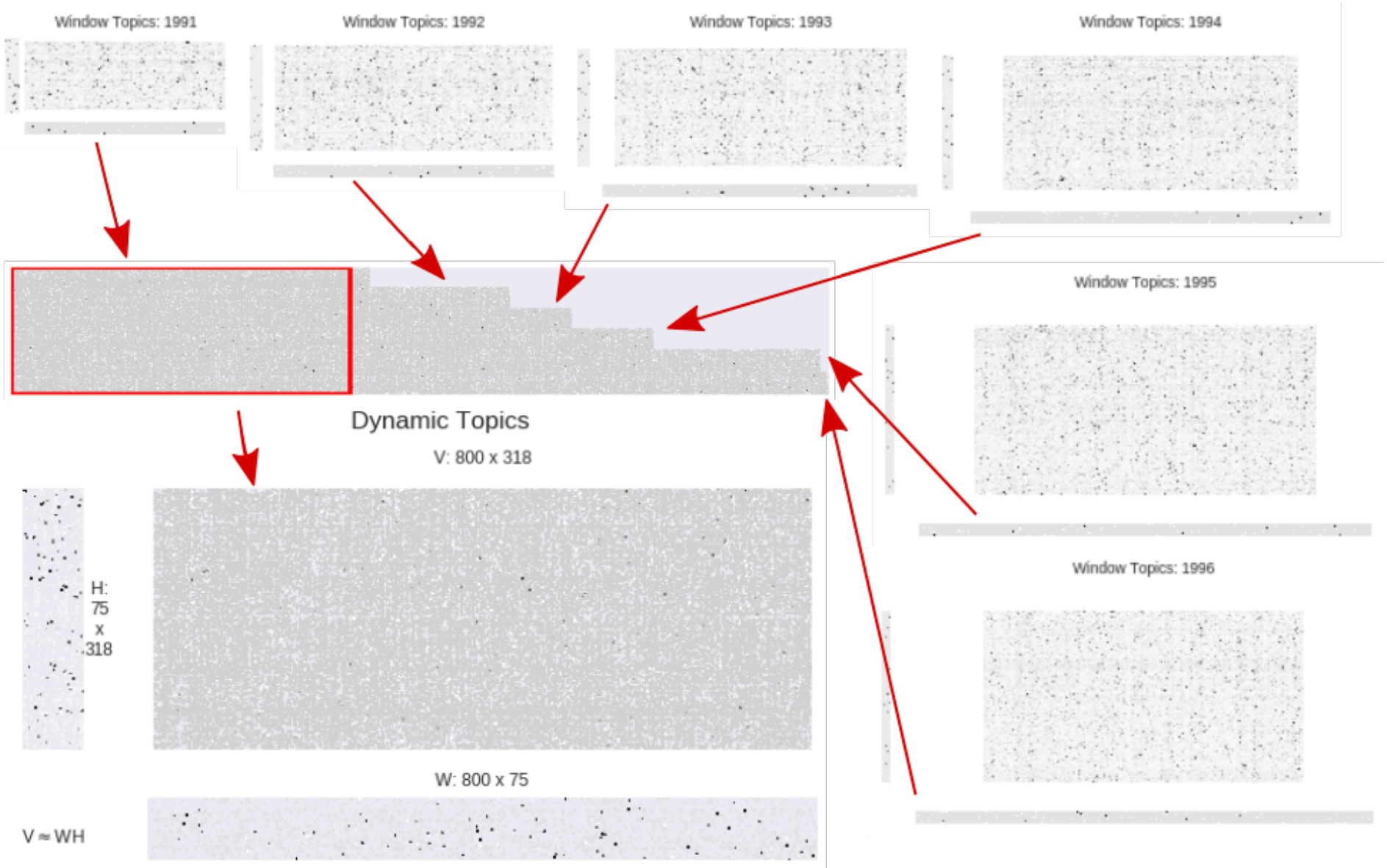
Dynamic NMF


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

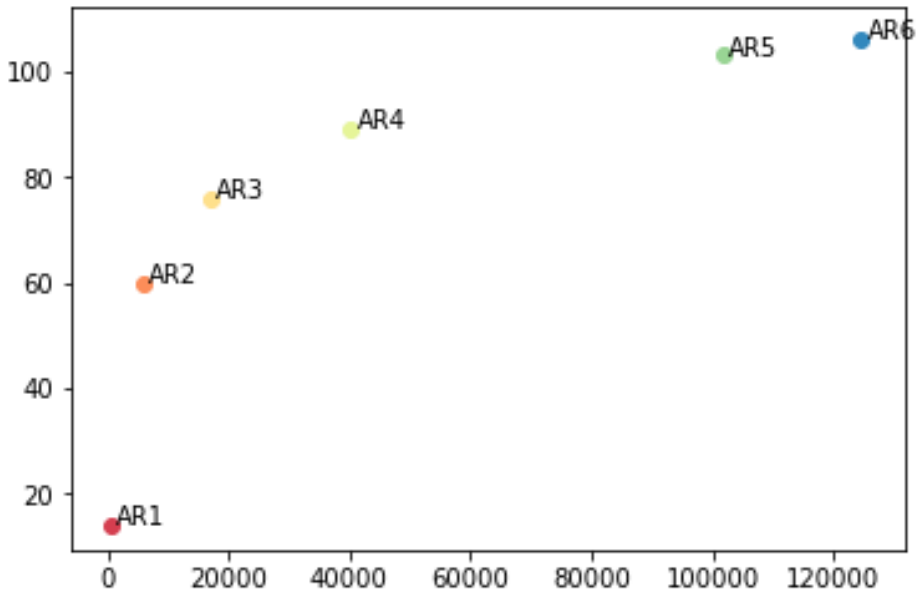
Dynamic NMF - application to climate change

- 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

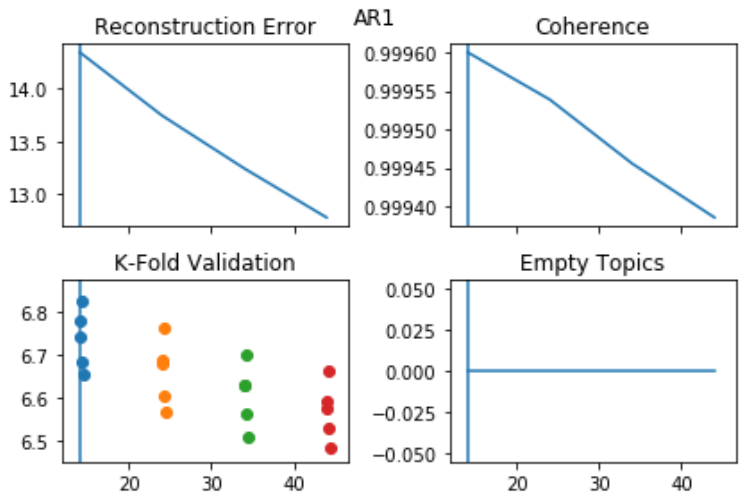
Dynamic NMF - number of topics

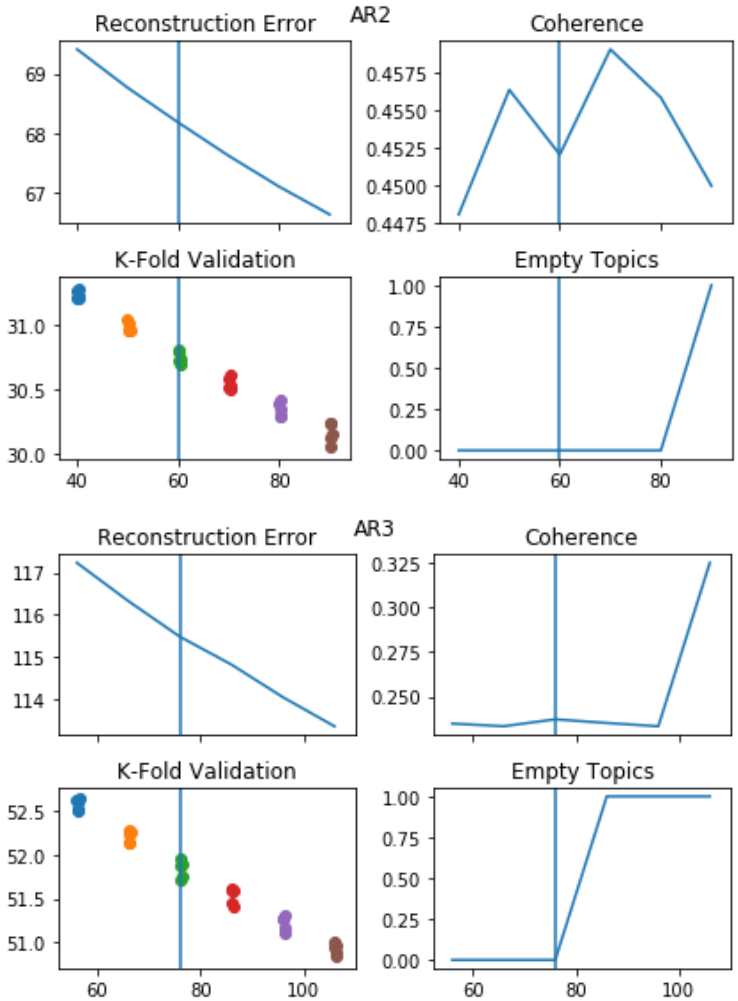
Human topic number criteria

- Intelligibility

Data-driven topic number criteria

- Reconstruction accuracy
- Predictive capacity





Results

Outline

Results

Outline

- Topography

Results

Outline

- Topography
- Structure

Results

Outline

- Topography
- Structure
- Development

Results

Outline

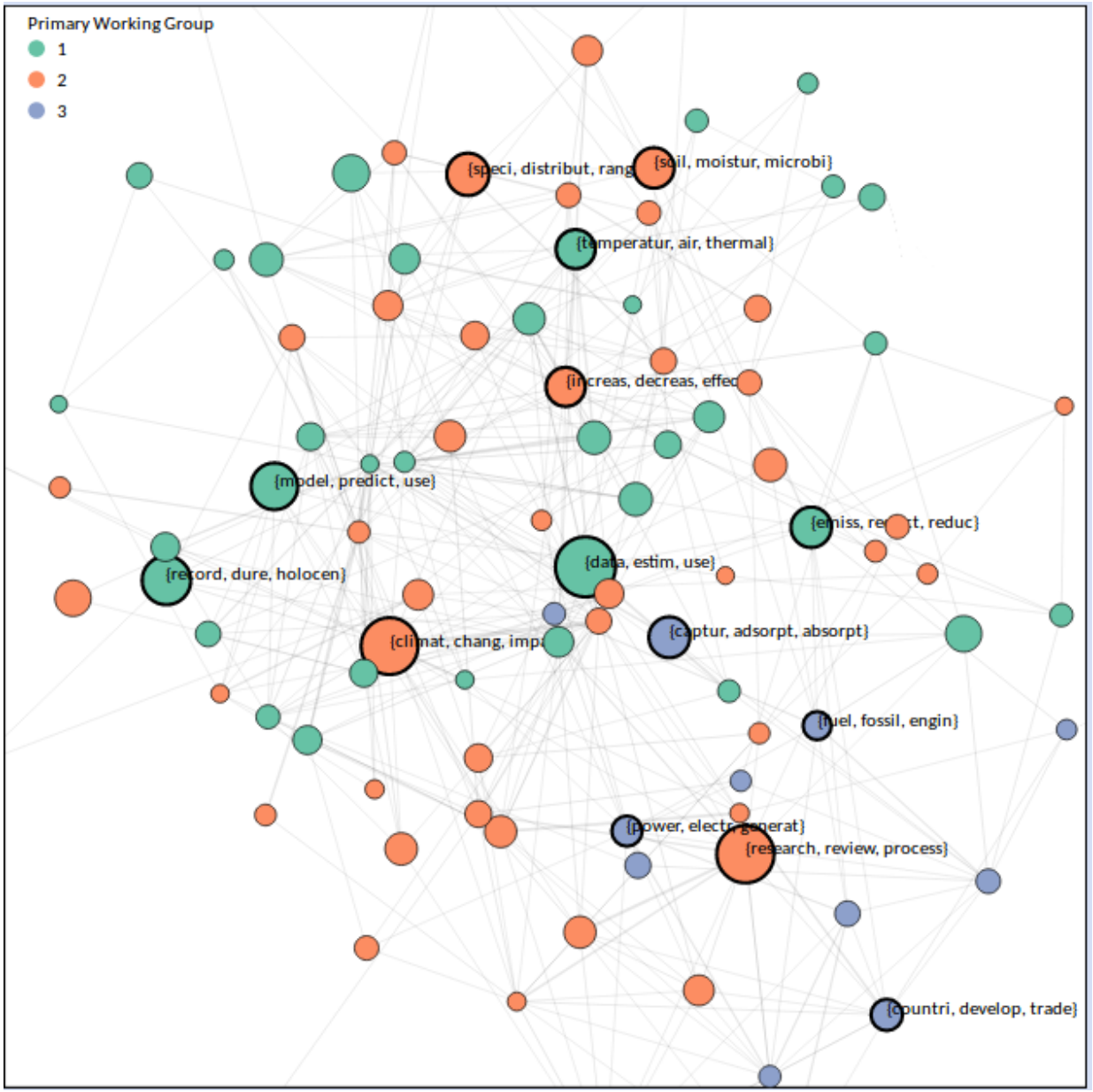
- Topography
- Structure
- Development
- Representation in past IPCC reports

Results

Outline

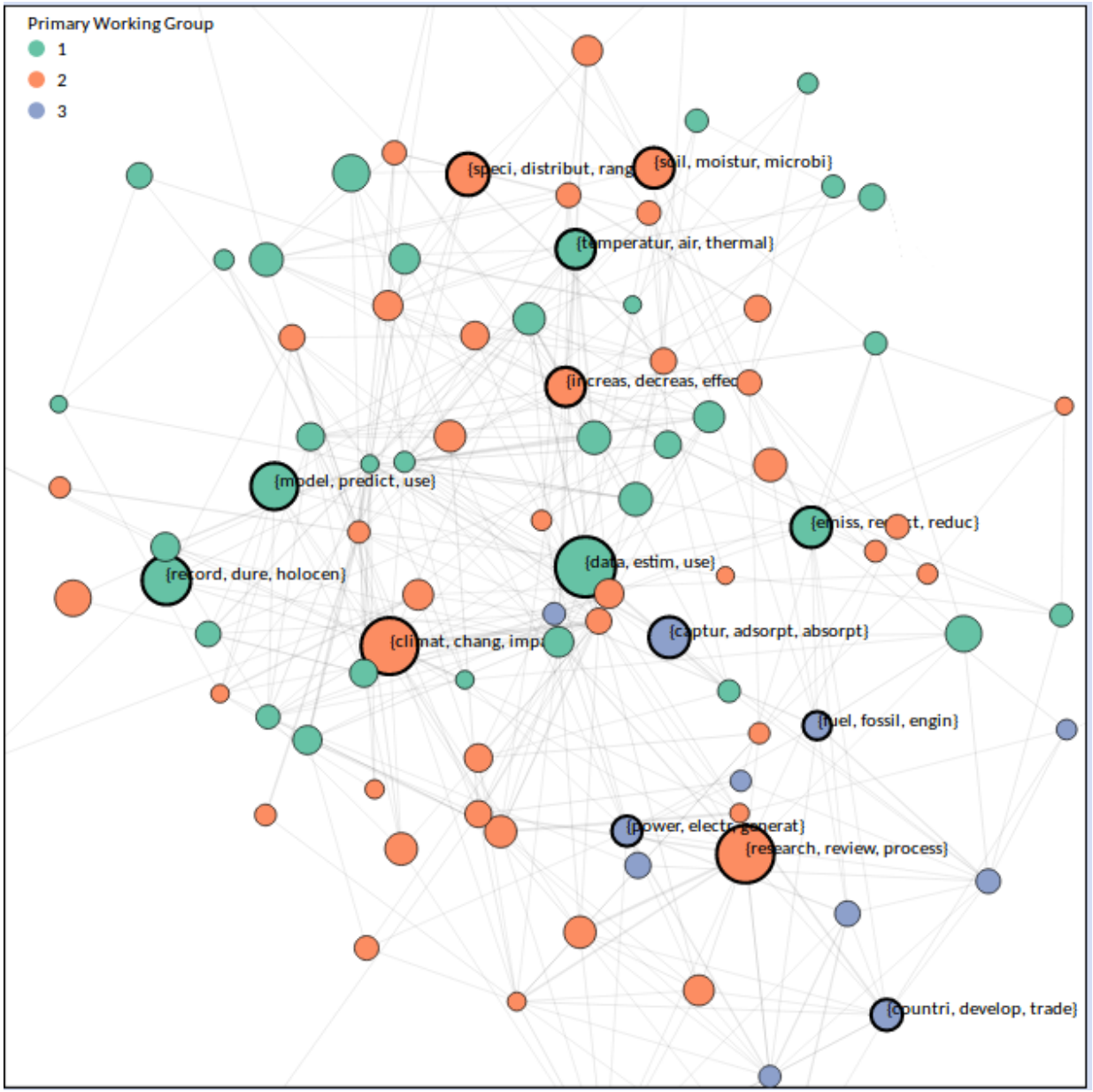
- Topography
- Structure
- Development
- Representation in past IPCC reports
- AR6 outlook

Results - structure



- Topics describe comprehensible themes in climate change research

Results - structure



- Topics describe comprehensible themes in climate change research
- Matching topics to the IPCC working group from which the majority of the topics are referenced in, a structure is generated based on topic-document correlations

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%

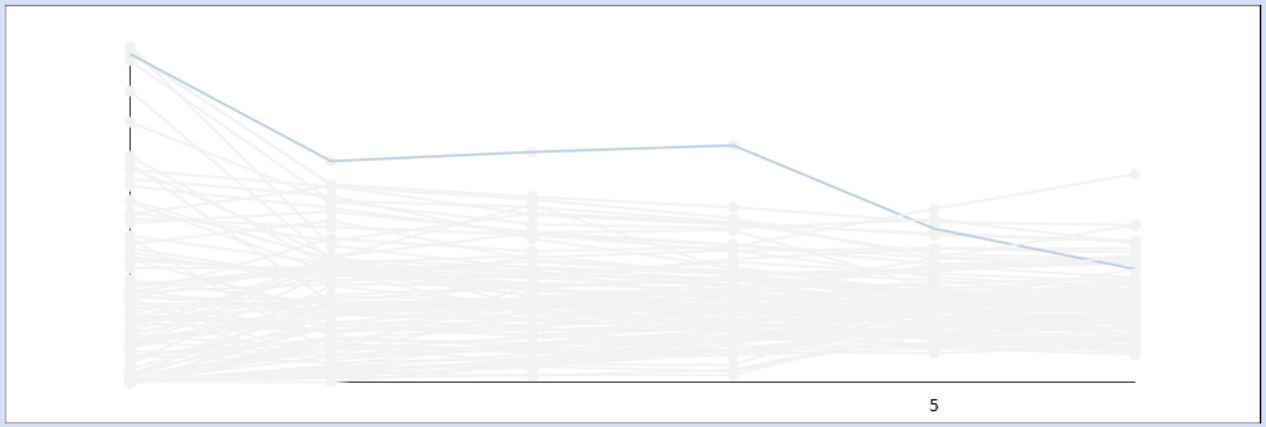
Results - Inter-working Group topics - WG 1 and 3

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%

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%

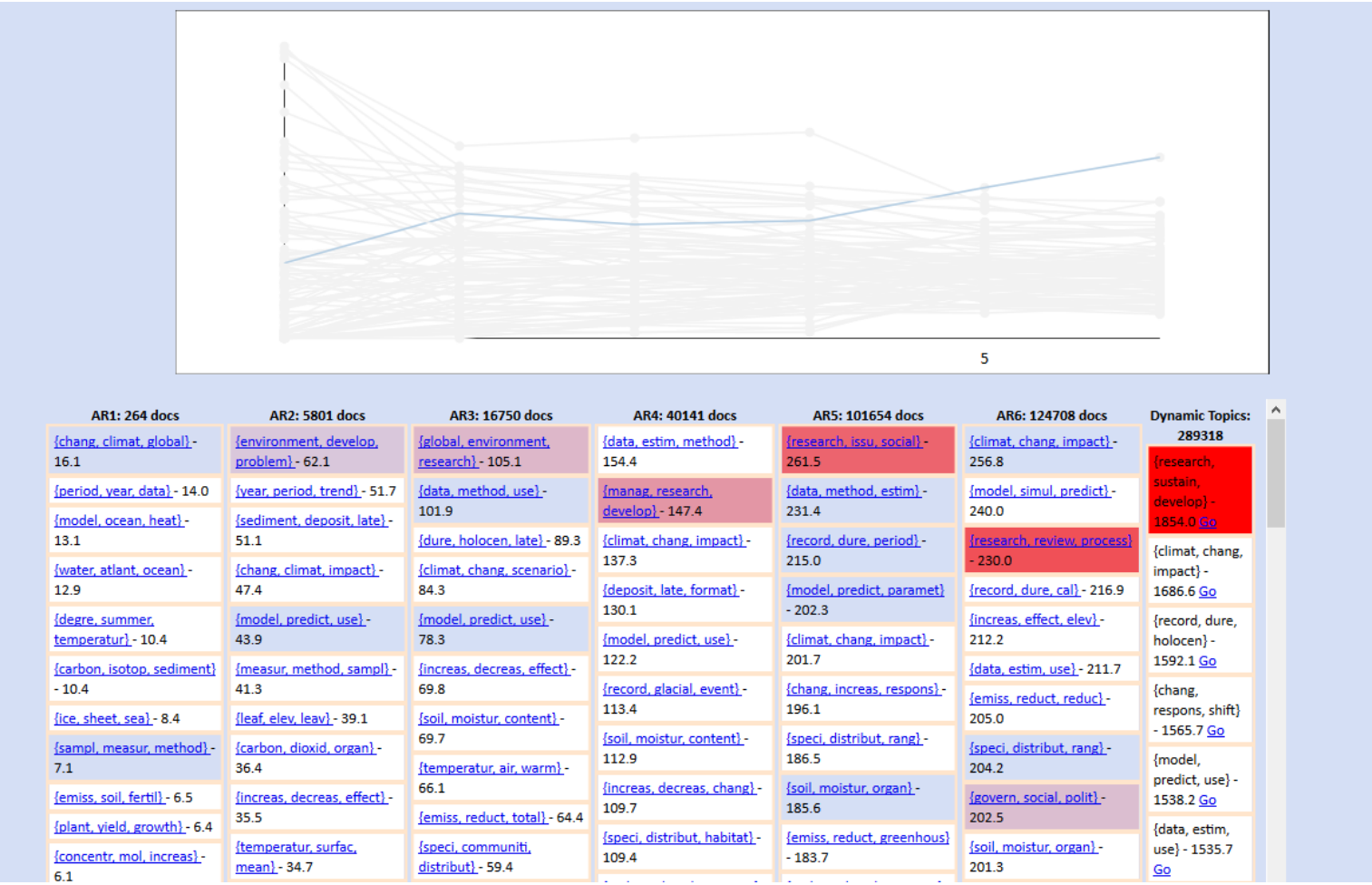
Results - Development



AR1: 264 docs	AR2: 5801 docs	AR3: 16750 docs	AR4: 40141 docs	AR5: 101654 docs	AR6: 124708 docs	Dynamic Topics: 289318
{chang, climat, global} - 16.1	{environment, develop, problem} - 62.1	{global, environment, research} - 105.1	{data, estim, method} - 154.4	{research, issu, social} - 261.5	{climat, chang, impact} - 256.8	{research, sustain, develop} - 1854.0 Go
{period, year, data} - 14.0	{year, period, trend} - 51.7	{data, method, use} - 101.9	{manag, research, develop} - 147.4	{data, method, estim} - 231.4	{model, simul, predict} - 240.0	{climat, chang, impact} - 1686.6 Go
{model, ocean, heat} - 13.1	{sediment, deposit, late} - 51.1	{dure, holocen, late} - 89.3	{climat, chang, impact} - 137.3	{record, dure, period} - 215.0	{research, review, process} - 230.0	{record, dure, holocen} - 1592.1 Go
{water, atlant, ocean} - 12.9	{chang, climat, impact} - 47.4	{climat, chang, scenario} - 84.3	{deposit, late, format} - 130.1	{model, predict, paramet} - 202.3	{record, dure, cal} - 216.9	{chang, respons, shift} - 1565.7 Go
{degre, summer, temperatur} - 10.4	{model, predict, use} - 43.9	{model, predict, use} - 78.3	{model, predict, use} - 122.2	{climat, chang, impact} - 201.7	{increas, effect, elev} - 212.2	{model, predict, use} - 1538.2 Go
{carbon, isotop, sediment} - 10.4	{measur, method, sampl} - 41.3	{increas, decreas, effect} - 69.8	{record, glacial, event} - 113.4	{chang, increas, respons} - 196.1	{data, estim, use} - 211.7	{data, estim, use} - 1535.7 Go
{ice, sheet, sea} - 8.4	{leaf, elev, leav} - 39.1	{soil, moistur, content} - 69.7	{soil, moistur, content} - 112.9	{speci, distribut, rang} - 186.5	{emiss, reduct, reduc} - 205.0	
{sampl, measur, method} - 7.1	{carbon, dioxid, organ} - 36.4	{temperatur, air, warm} - 66.1	{increas, decreas, chang} - 109.7	{soil, moistur, organ} - 185.6	{speci, distribut, rang} - 204.2	
{emiss, soil, fertil} - 6.5	{increas, decreas, effect} - 35.5	{emiss, reduct, total} - 64.4	{speci, distribut, habitat} - 109.4	{emiss, reduct, greenhous} - 183.7	{govern, social, polit} - 202.5	
{plant, yield, growth} - 6.4	{temperatur, surfac, mean} - 34.7	{speci, communiti, distribut} - 59.4	{carbon, dioxid, sequestr} -	{carbon, dioxid, sequestr} -	{soil, moistur, organ} - 201.3	
{concentr, mol, increas} - 6.1	{speci, popul, distribut} -	{record, core, isotop} -			{carbon, dioxid, sequestr} -	{emiss, reduct

Basic climate science topics are not as prominent as they were previously

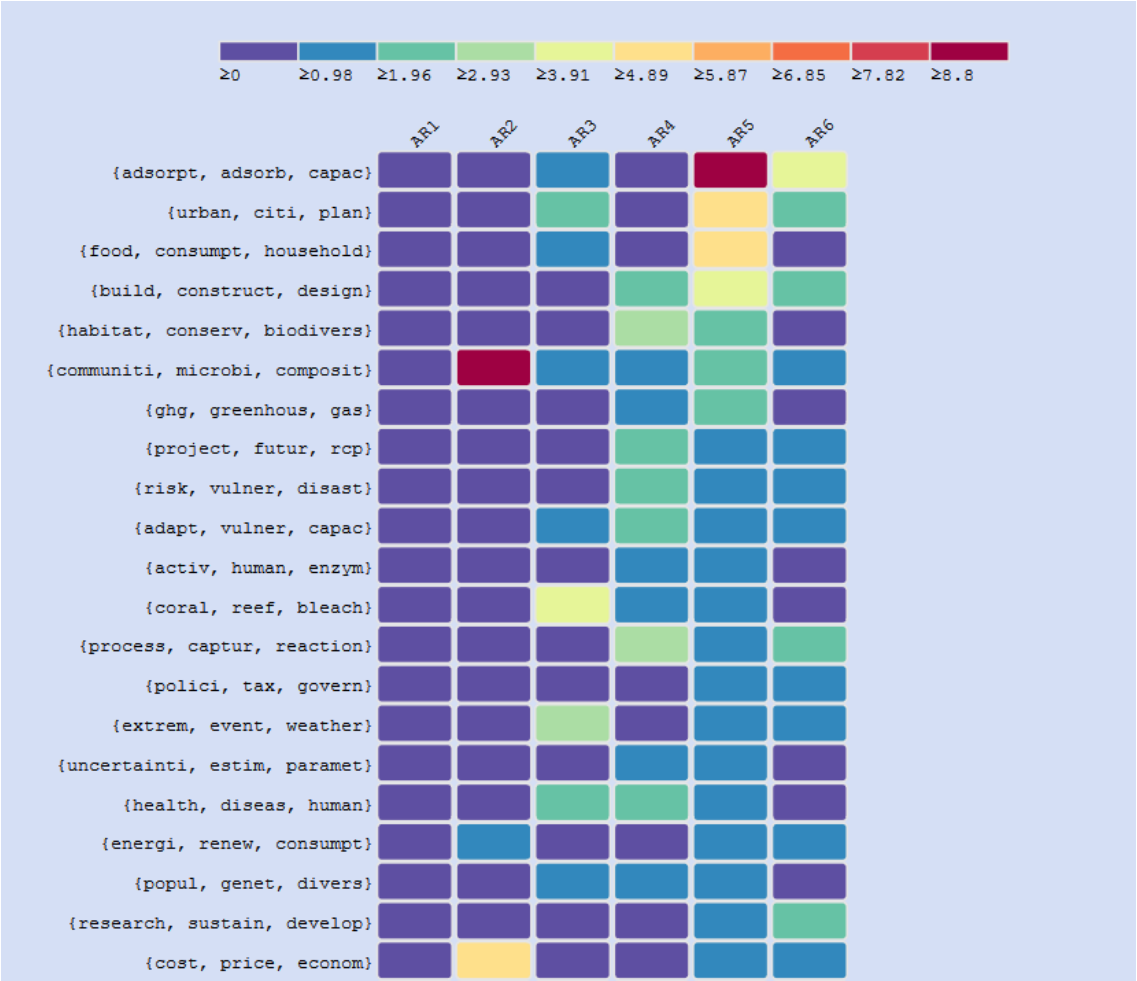
Results - Development



Sustainable development, and research agendas are more prominent

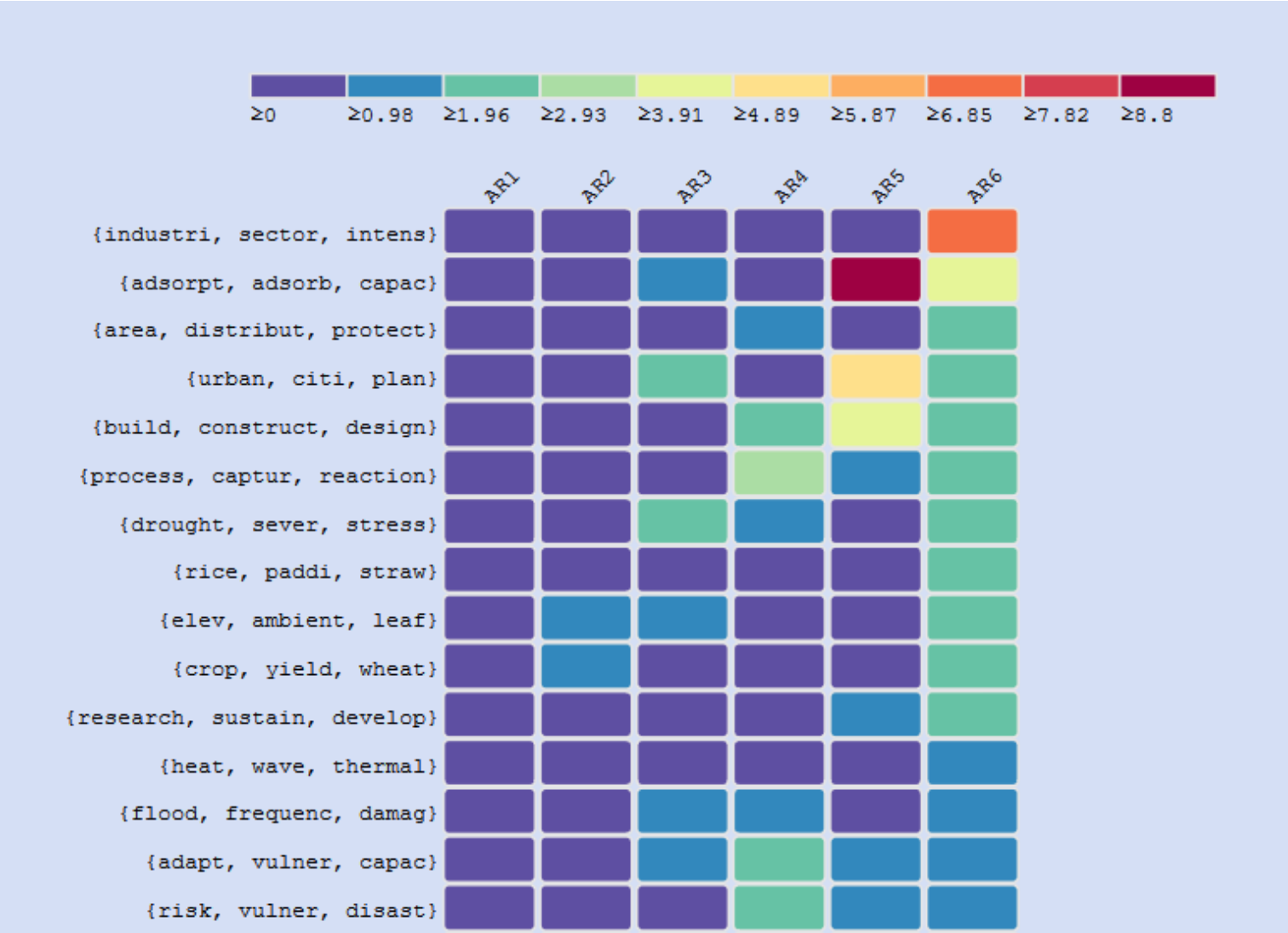
The documents that underpin this growth

Results - Development



Fast growing topics in AR5 were on urban systems, negative emissions, buildings, consumption, biodiversity and risks

Results - Development



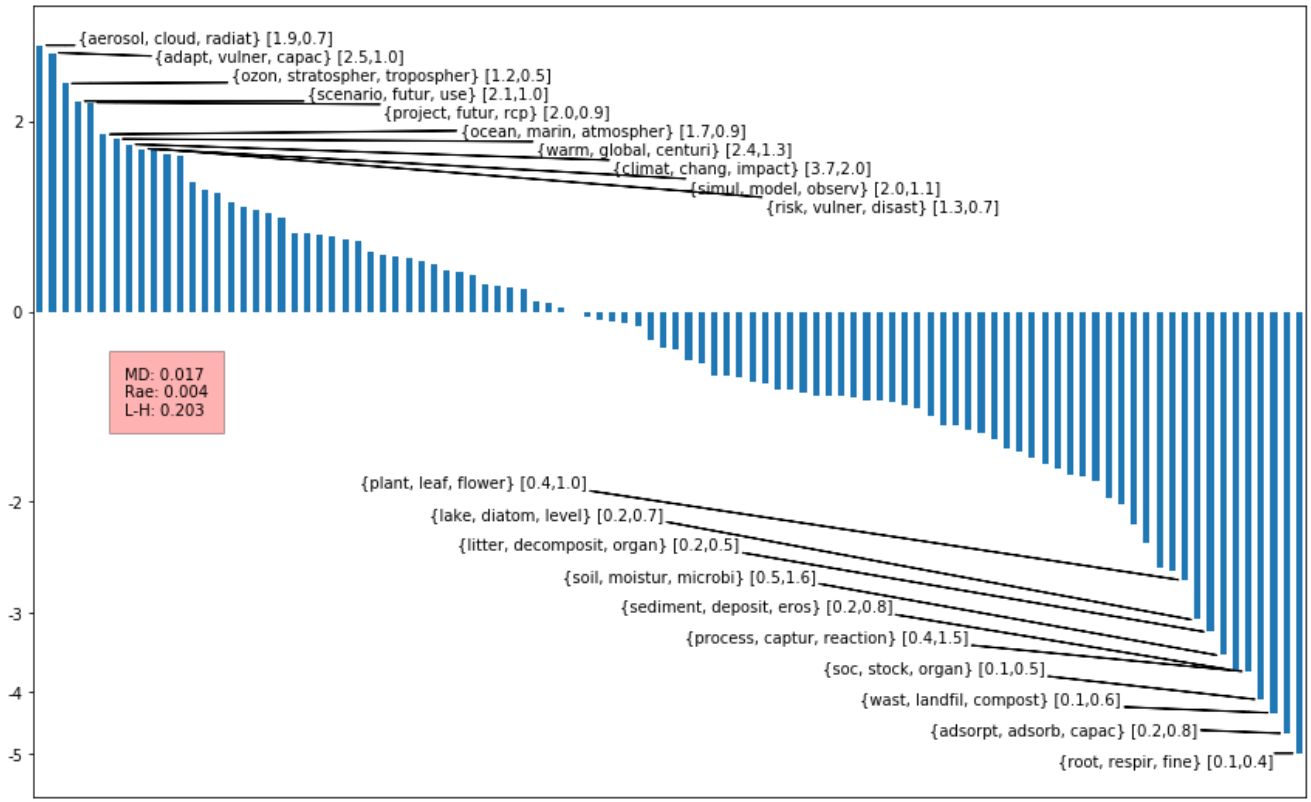
Negative emissions topics continue to grow,

Results - topic representation in IPCC reports

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

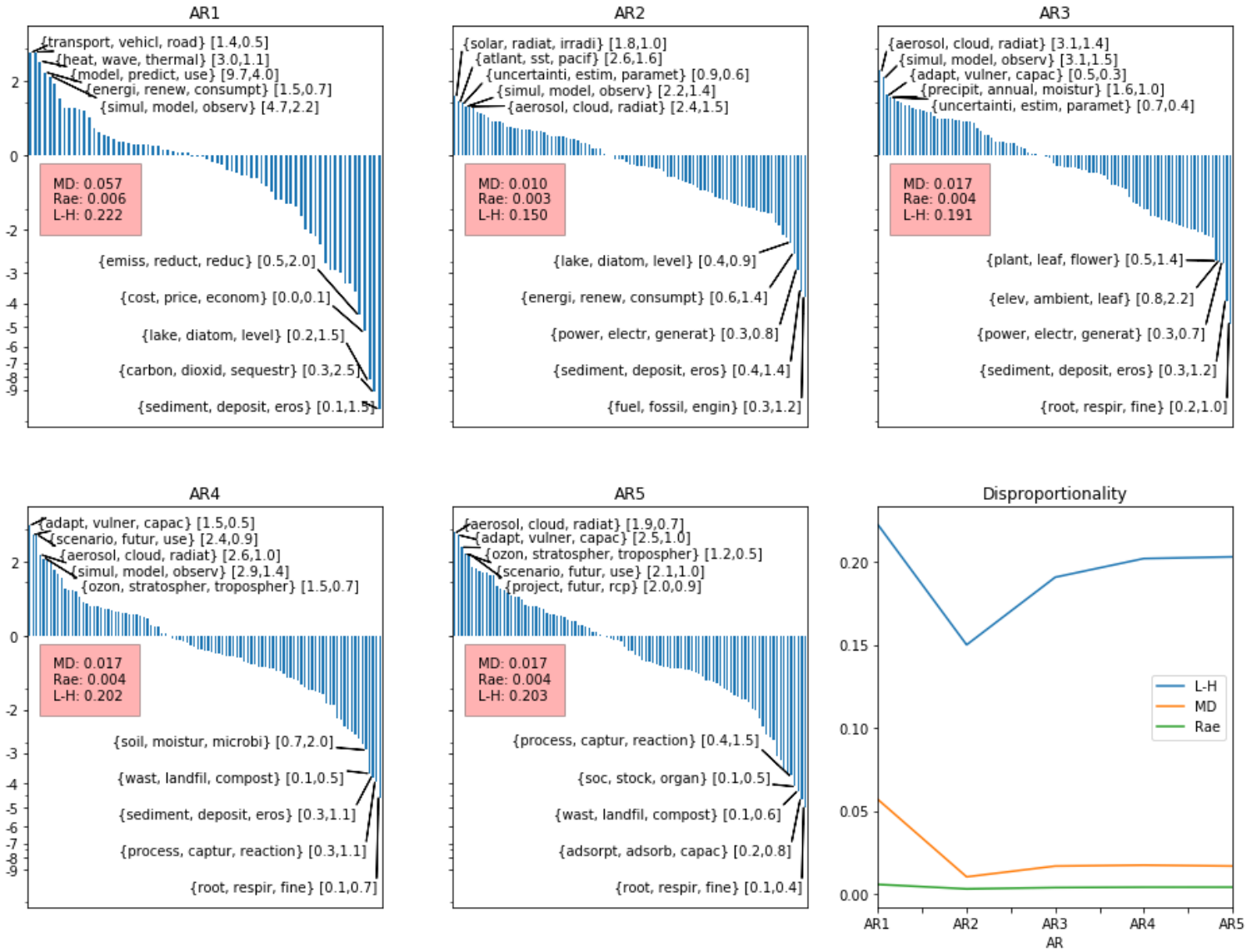
We get a measure of proportionality between two distributions by dividing the each topics share in the IPCC sample by its share in the whole corpus

Results - topic representation in IPCC reports

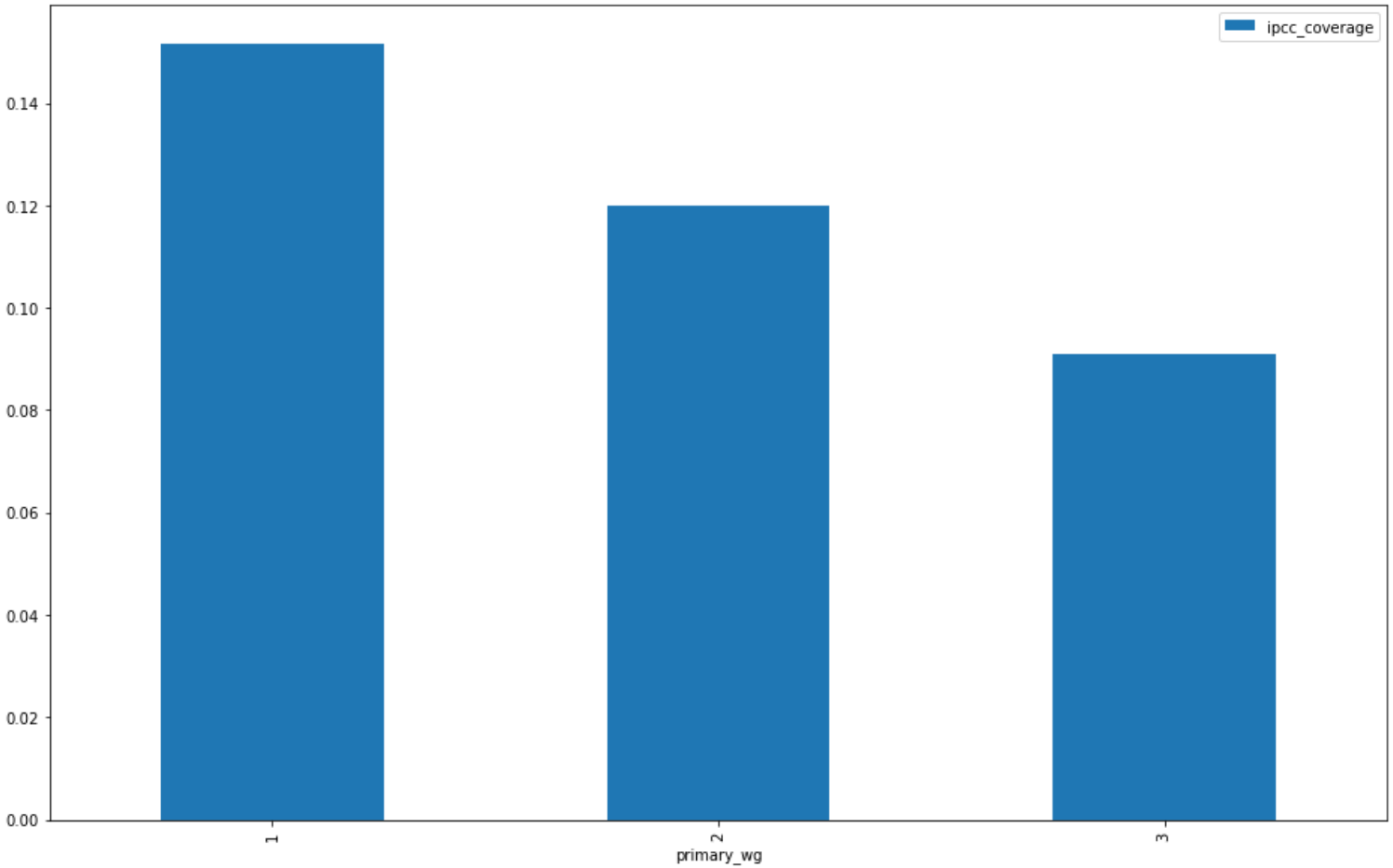


- The physical science aspects of climate change, as well topics on impacts, adaptation and scenarios are well covered by the IPCC
- Topics on specific technological solutions (particularly NETs), as well as soils and plants are less well represented

Results - topic representation in IPCC reports



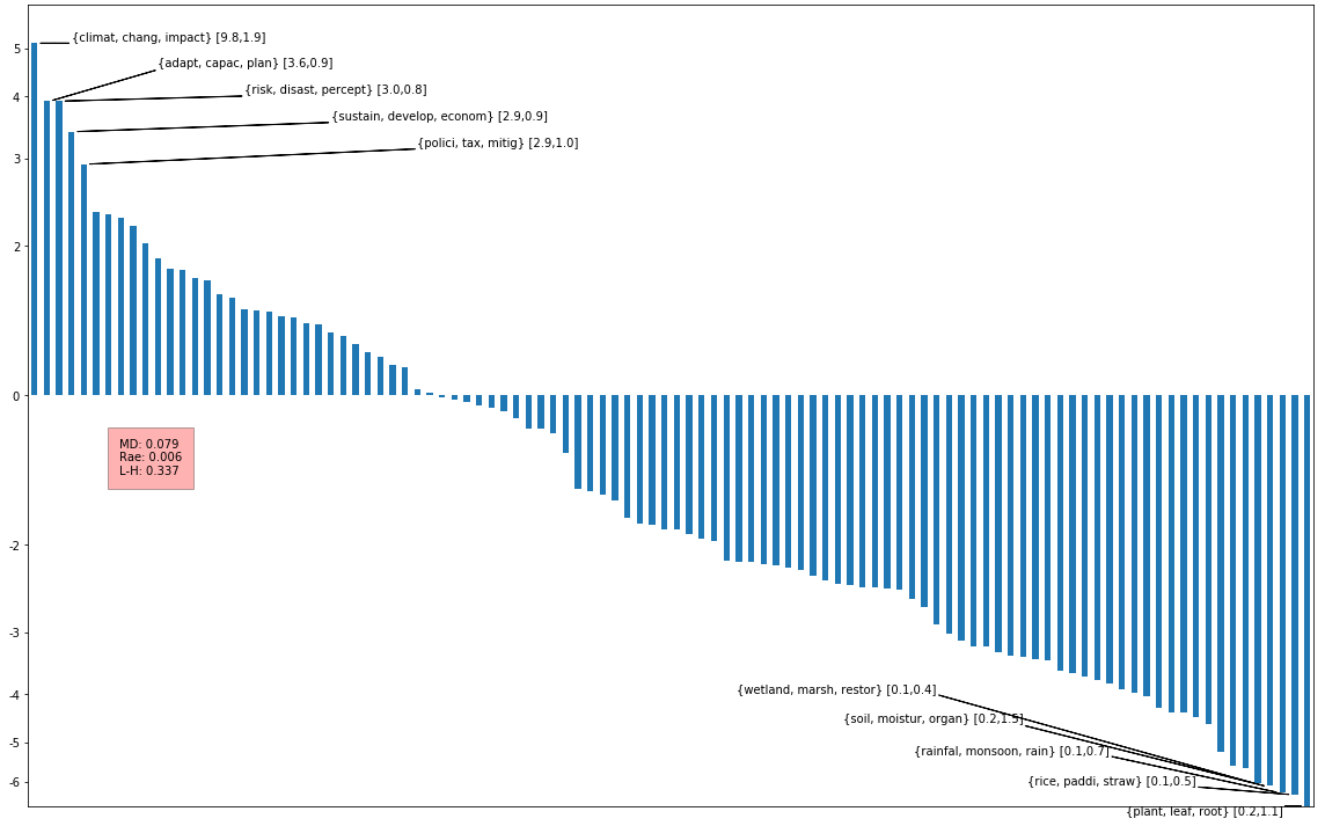
Results - topic representation in IPCC reports



Results - AR6 outlook

Outlines for the sixth assessment report have already been published.
The words in each chapter were matched to the Topic-Term matrix generated by the model to calculate a [simplified] topic score for each chapter.
These scores were summed and were compared to the corpus in general in the same way as with past reports

Results - topic representation in IPCC reports



- 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 and plants are less well represented

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

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.