A Topography of Climate Change Research

Max Callaghan

July 26, 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`

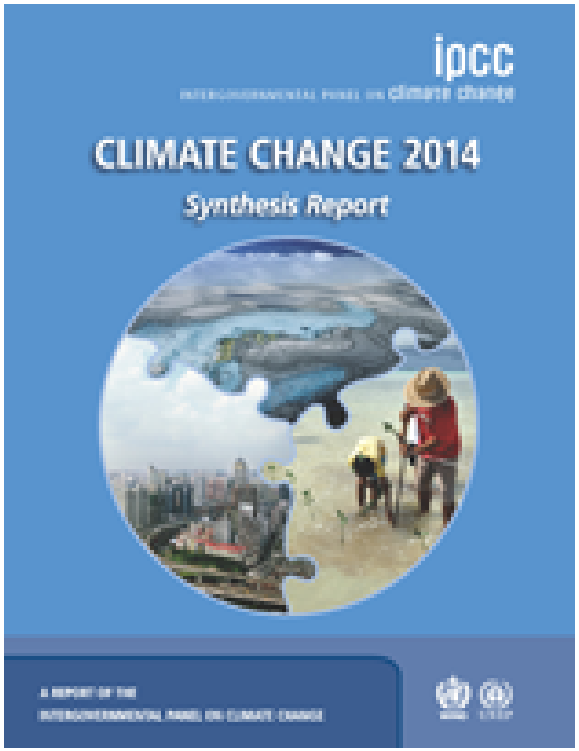
- Topography is a description of a landscape

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`

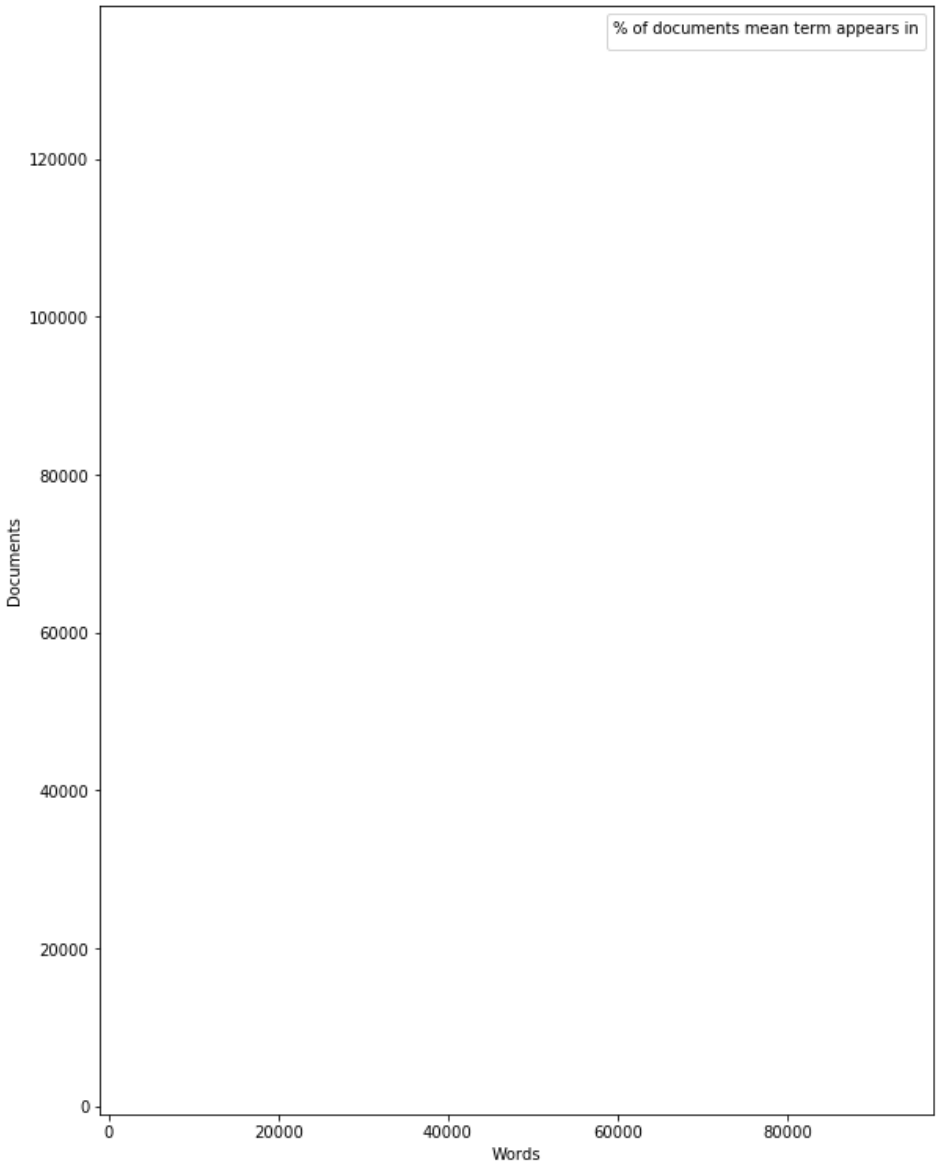
- Topography is a description of a landscape
- Topics (from the Greek τοπος, place) can describe the features of a body of text

Context



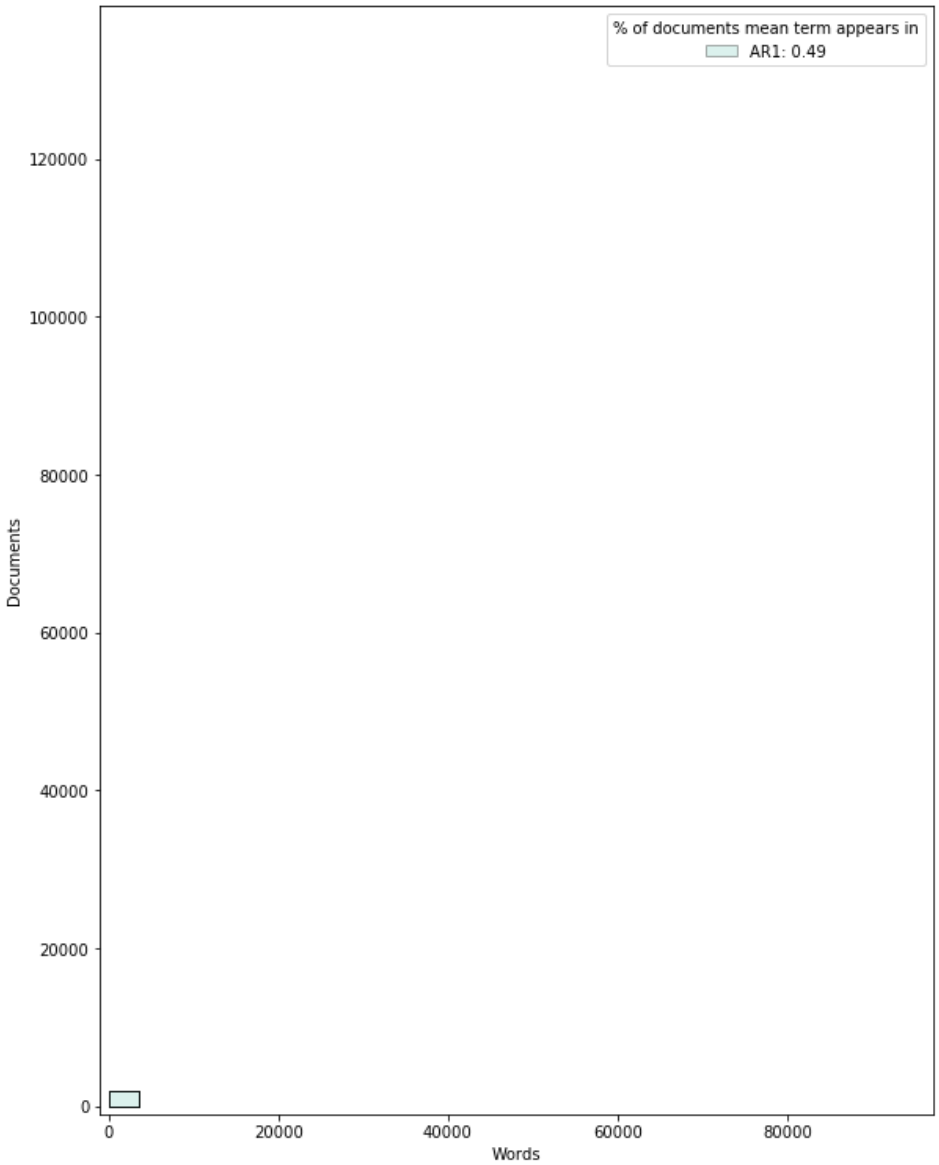
- To contribute evidence-based policy-making on climate change, the IPCC aims to *comprehensively* assess scientific literature on climate change
- These assessments should be aim to balance legitimacy, credibility and relevance (Cash and Clark, 2001)

Motivation - Big Literature



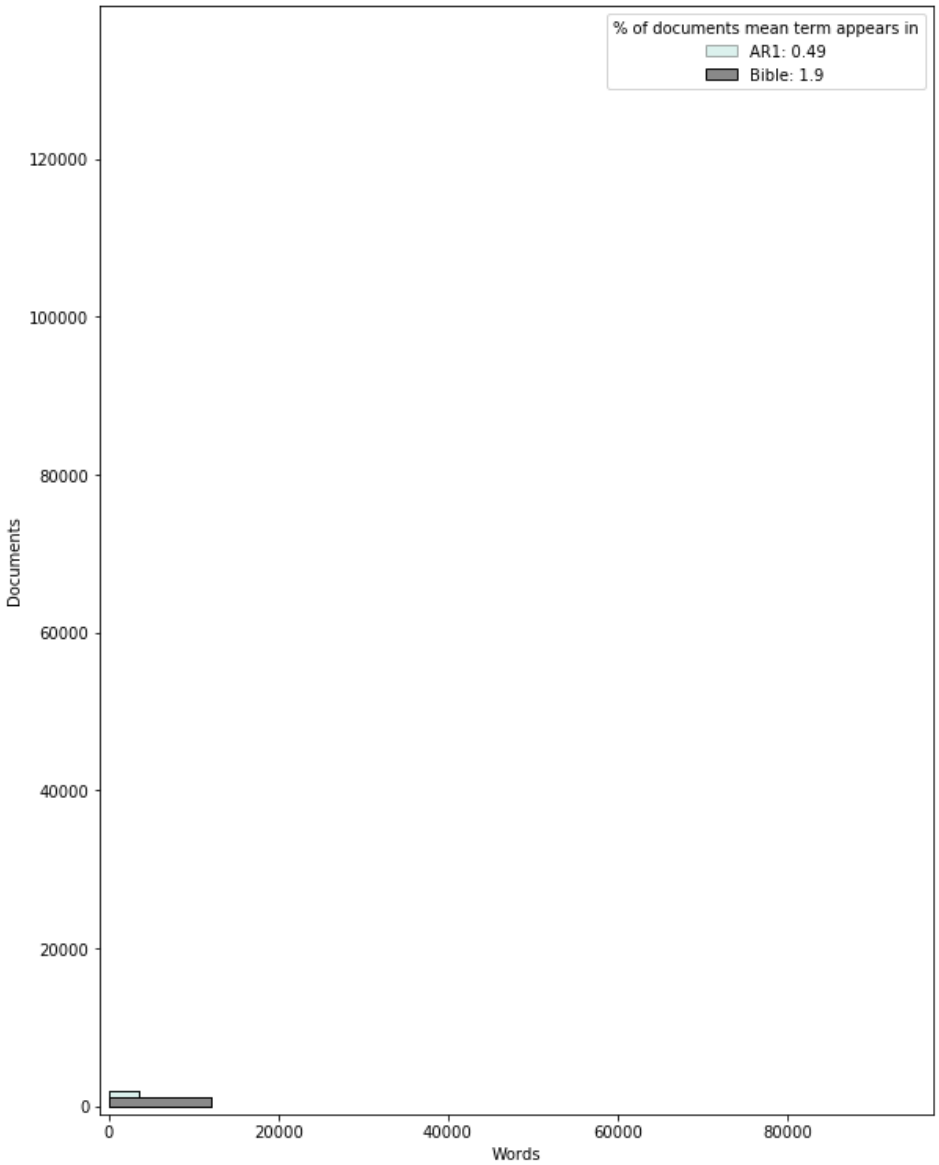
A matrix of documents x words

Motivation - Big Literature



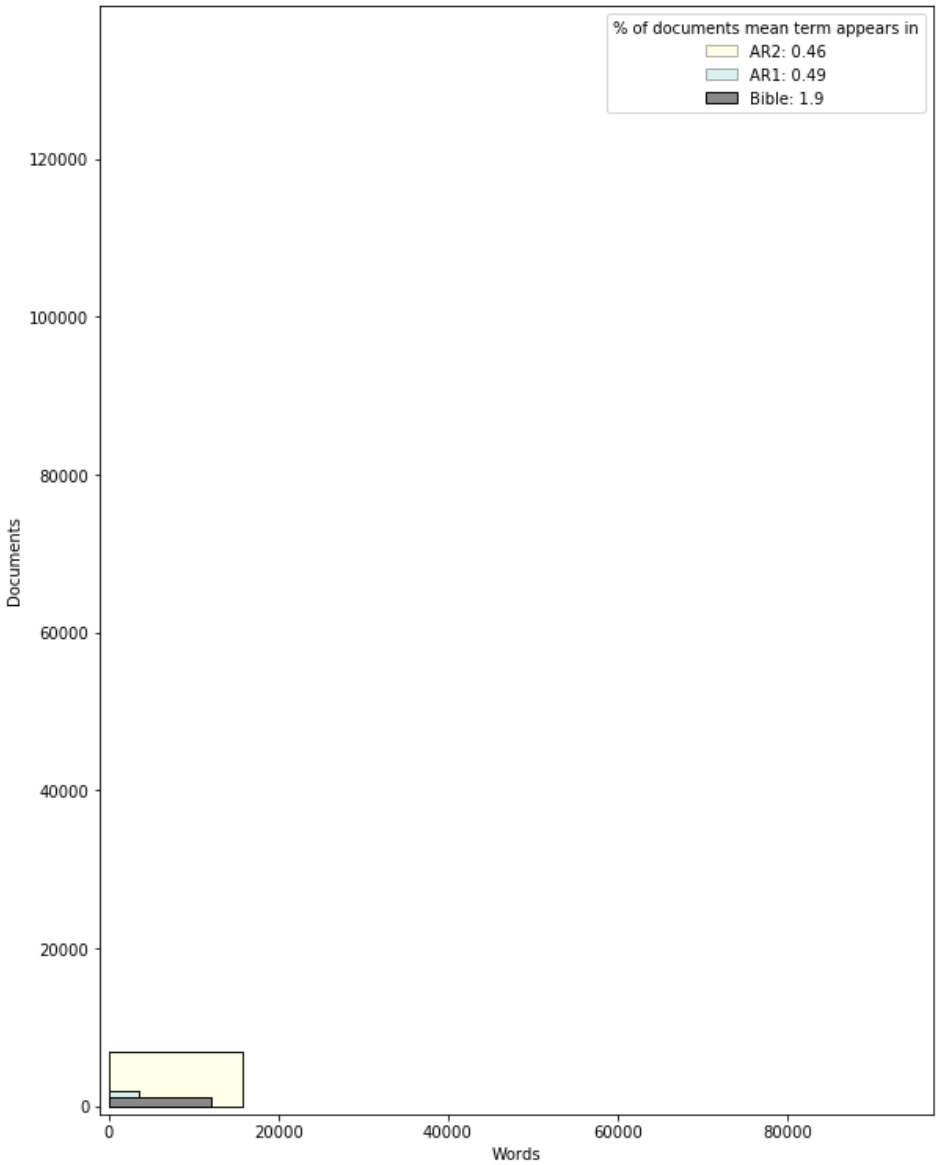
AR1: 1,848 documents x 3,528 words

Motivation - Big Literature



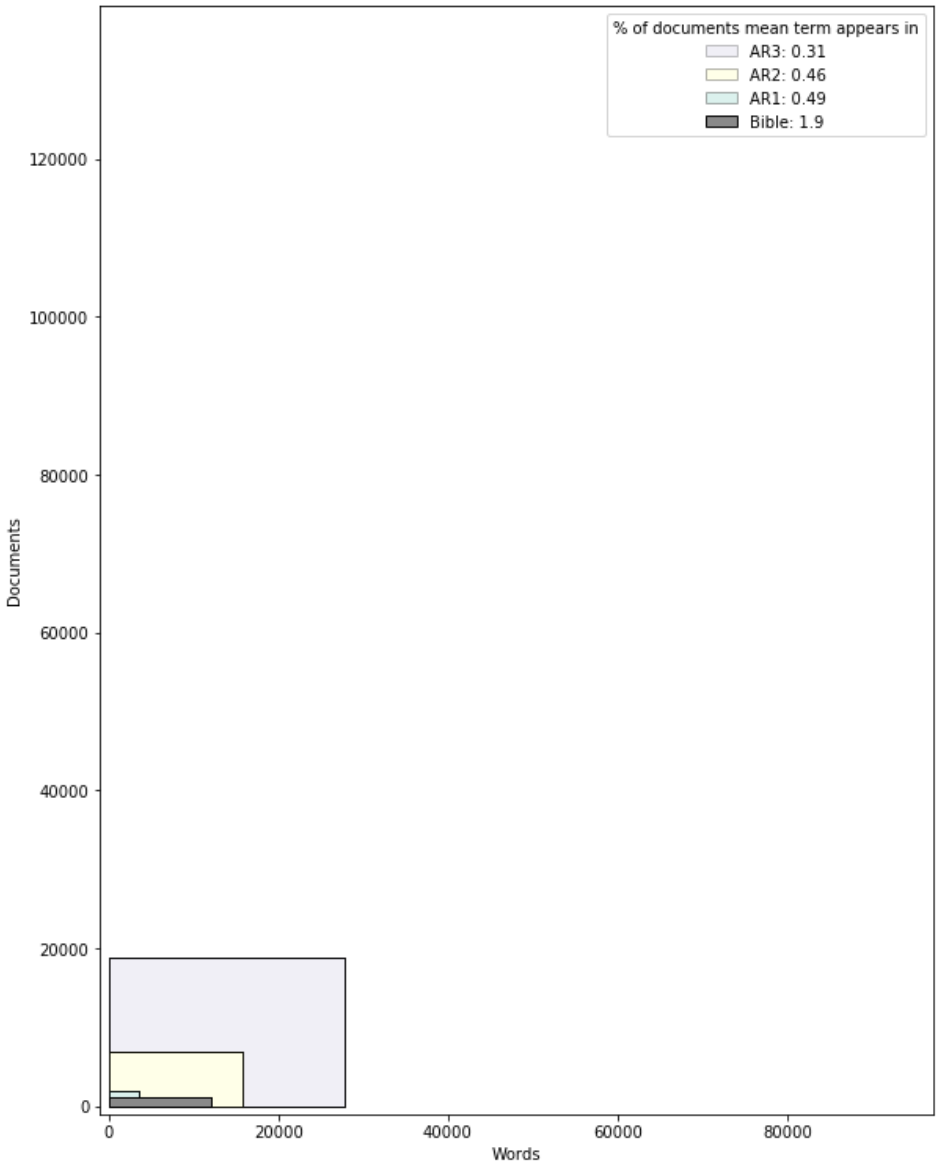
The Luther Bible: 1,189 documents (chapters) x 11,973 words

Motivation - Big Literature



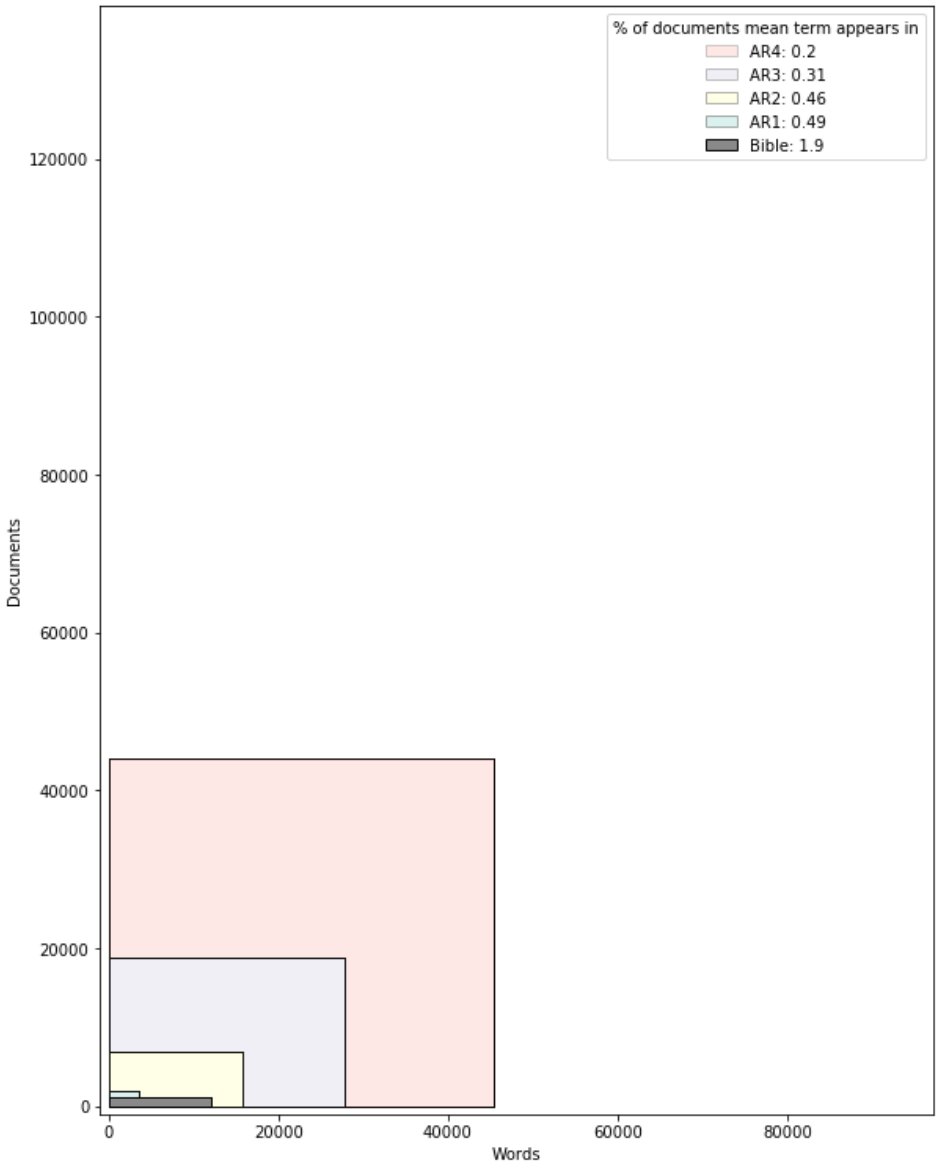
AR2: 6,941 documents x 15,781 words

Motivation - Big Literature



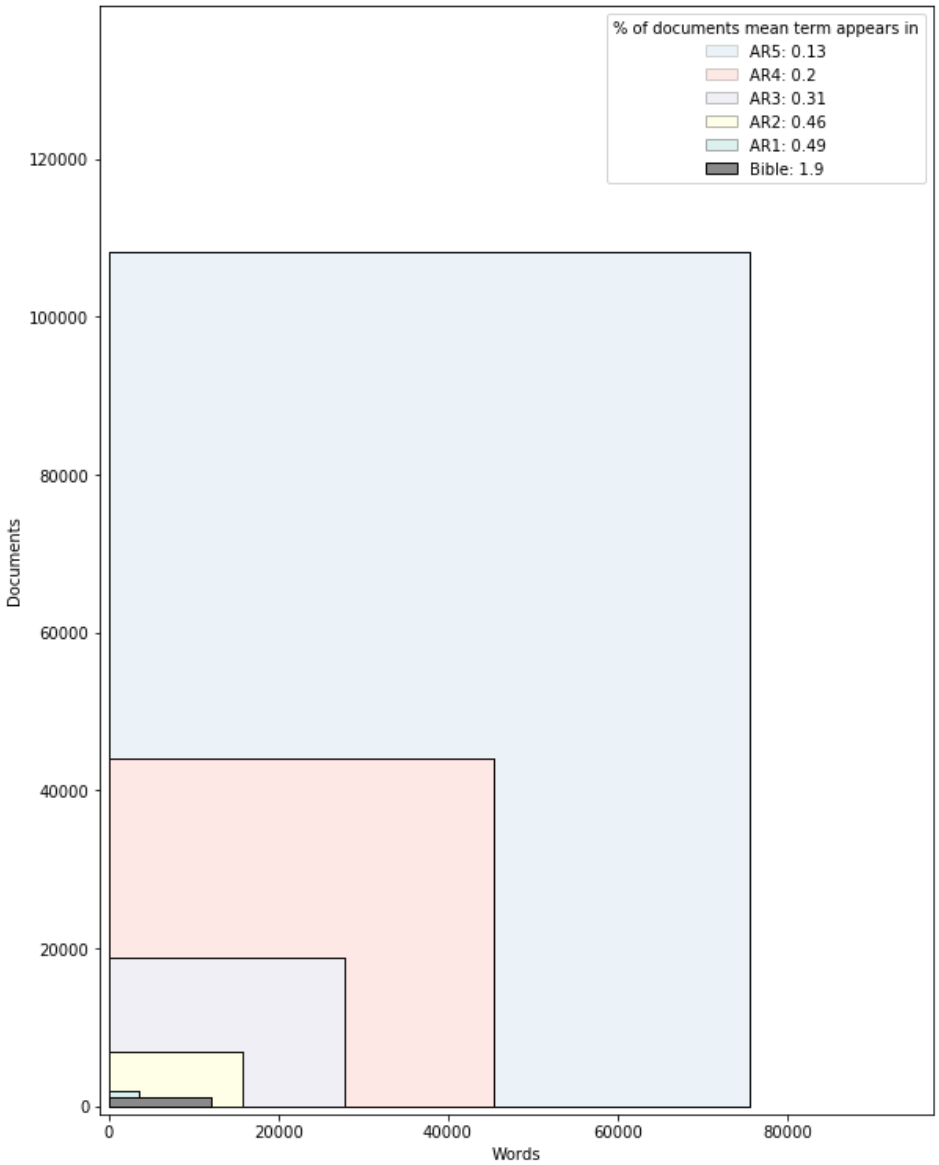
AR3: 18,728 documents x 27,730 words

Motivation - Big Literature



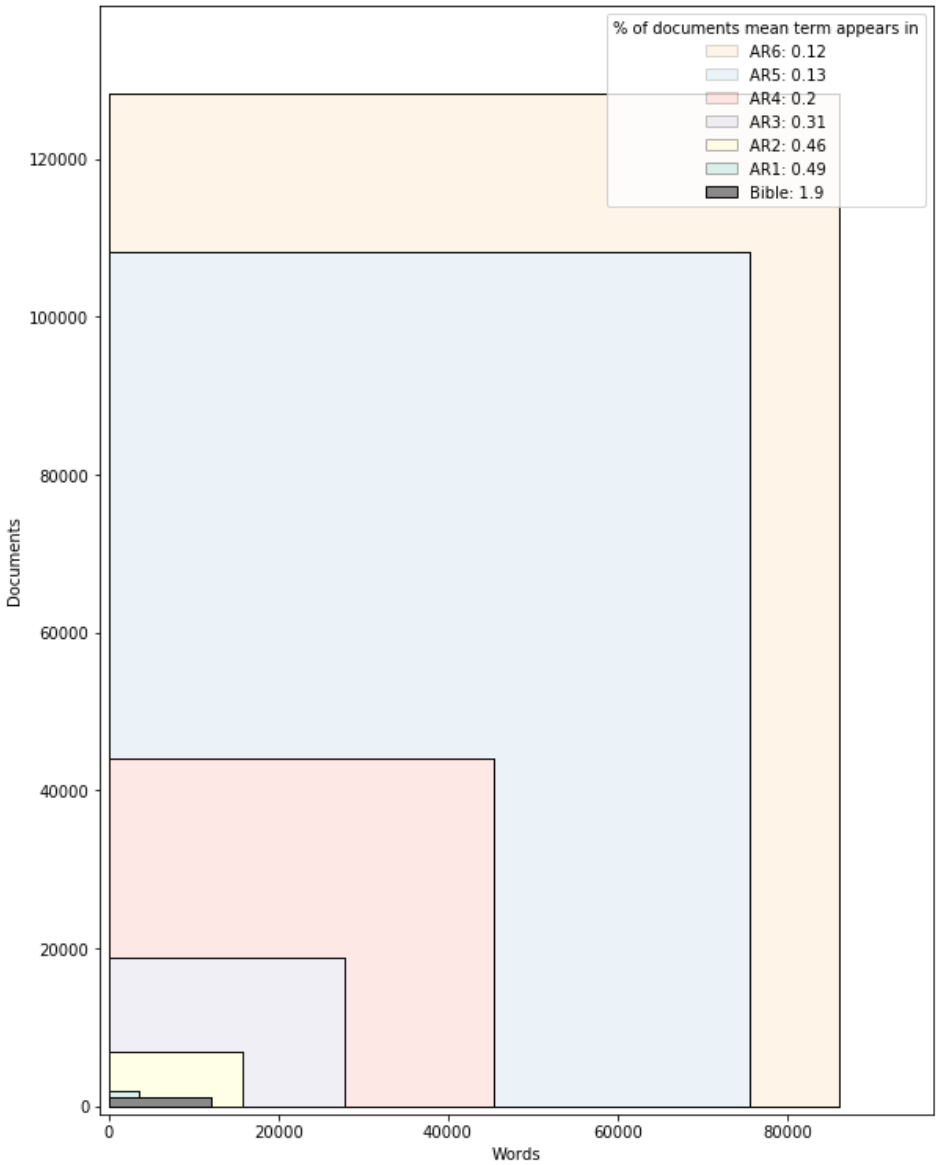
AR4: 44,000 documents x 45,388 words

Motivation - Big Literature



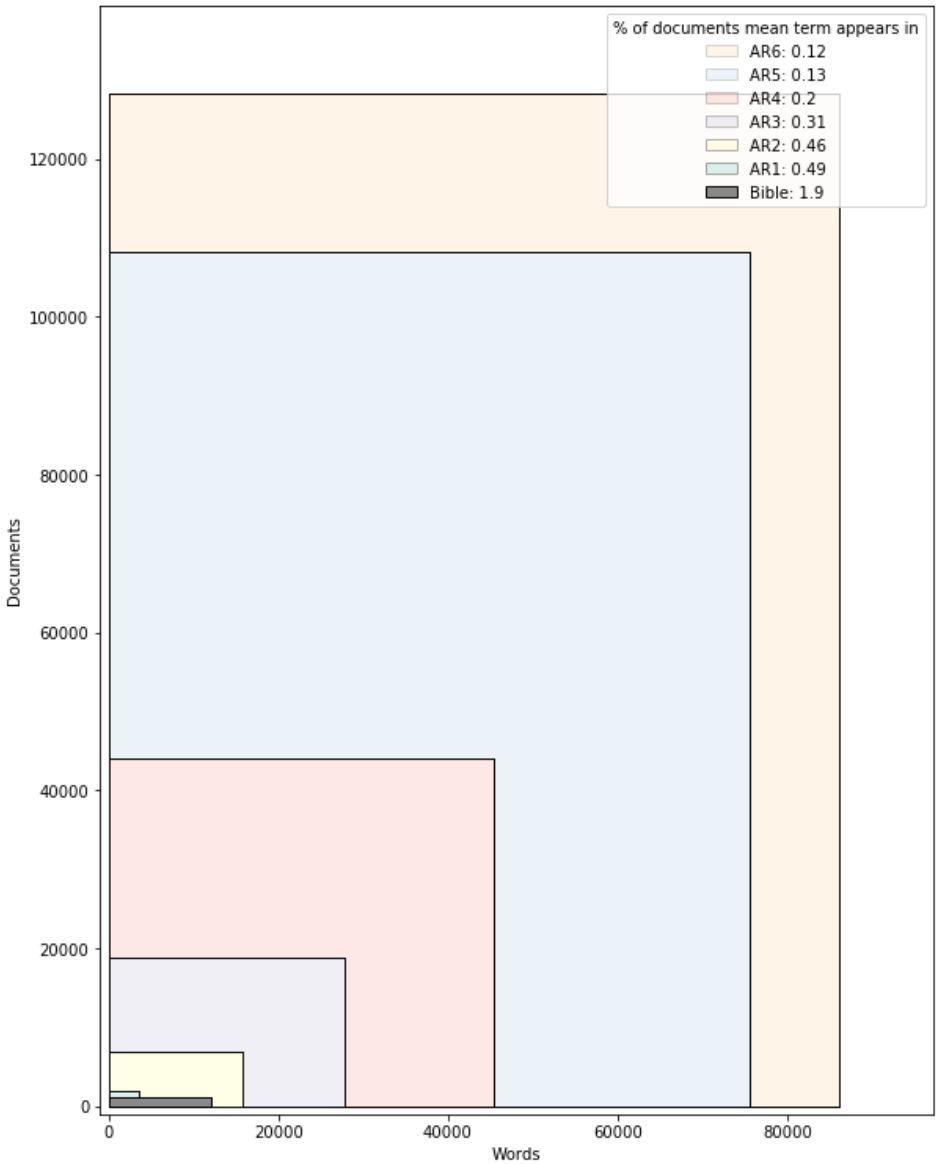
AR5: 108,277 documents x 75,553 words

Motivation - Big Literature



AR6: 128,357 documents x 86,149 words

Motivation - Big Literature

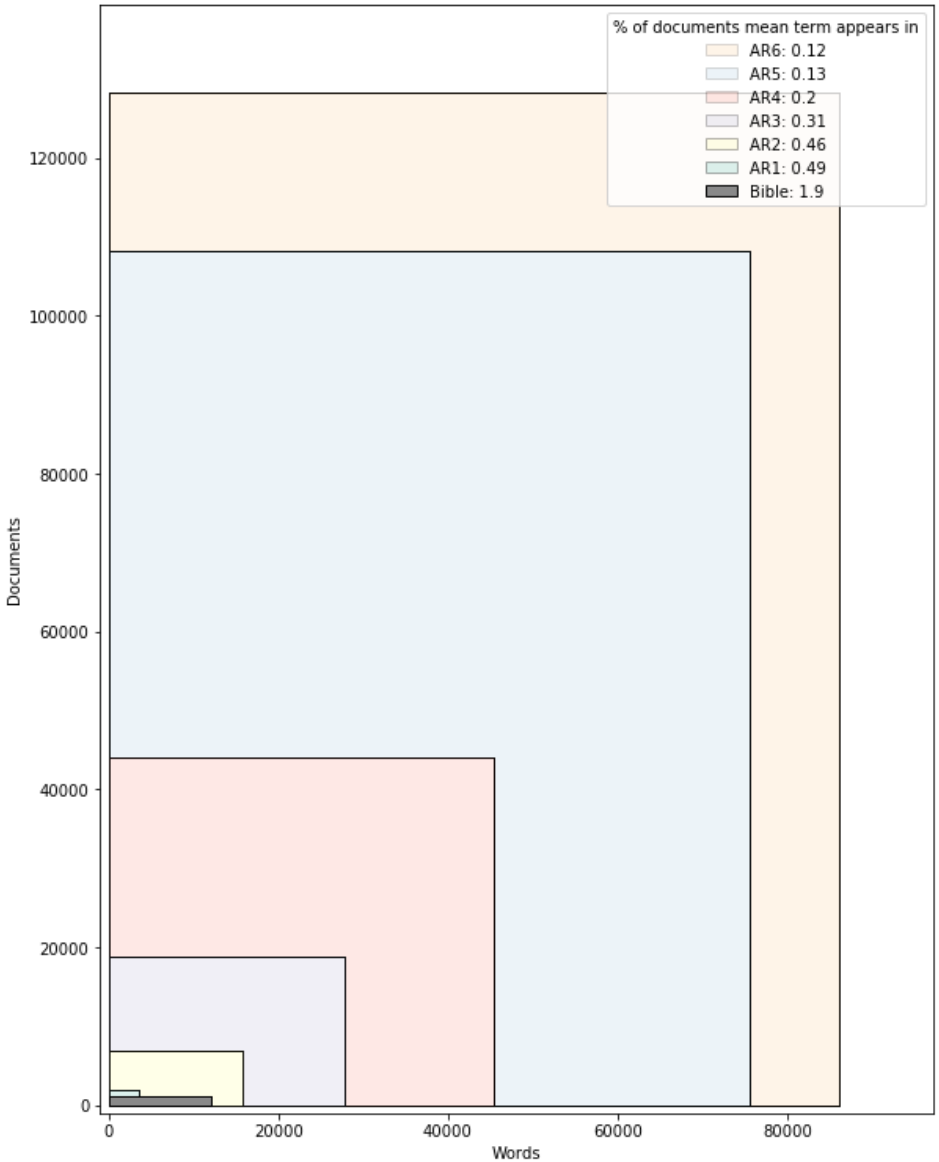


AR6: 128,357 documents x 86,149 words

- Comprehensive, credible and relevant assessments become more challenging as the literature grows (Minx et al., 2017)

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

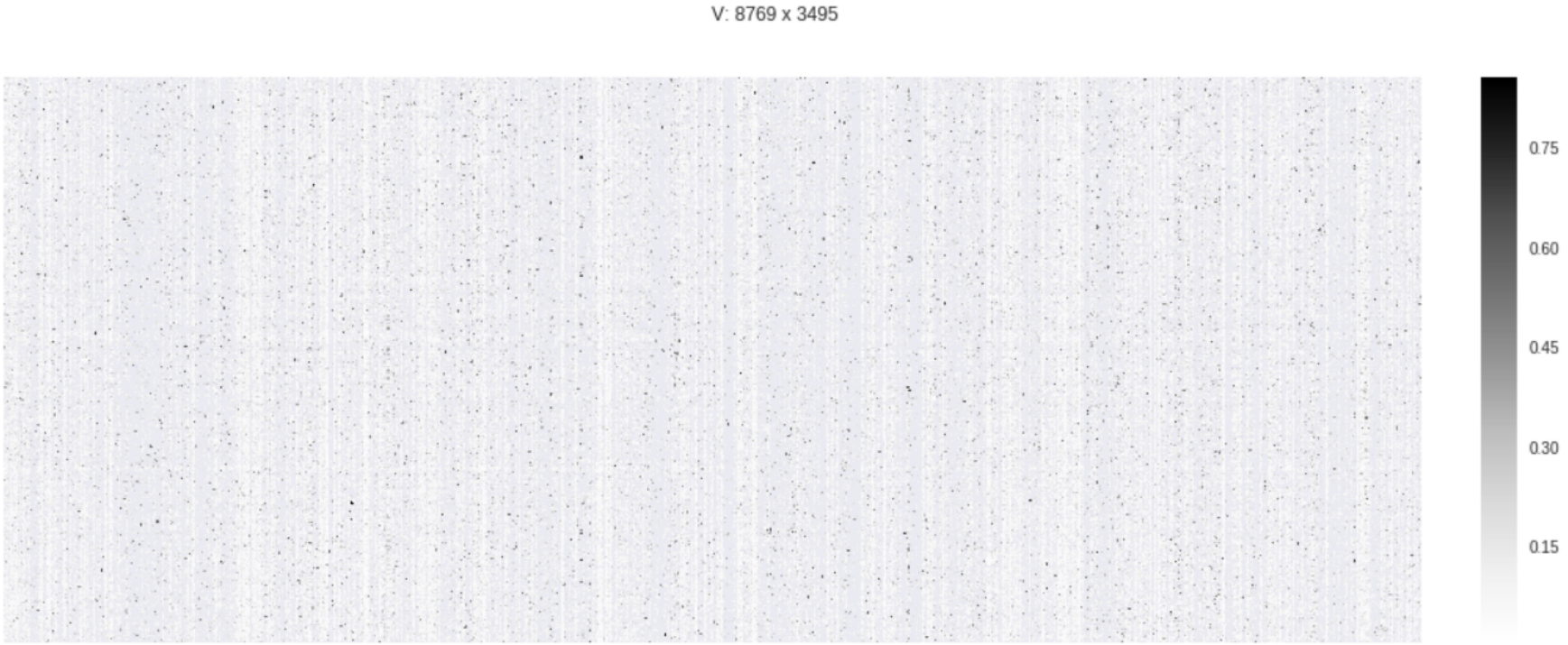


Figure: A topic model of 3495 documents on climate change from the year 2000

Approach - Words, words, words

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

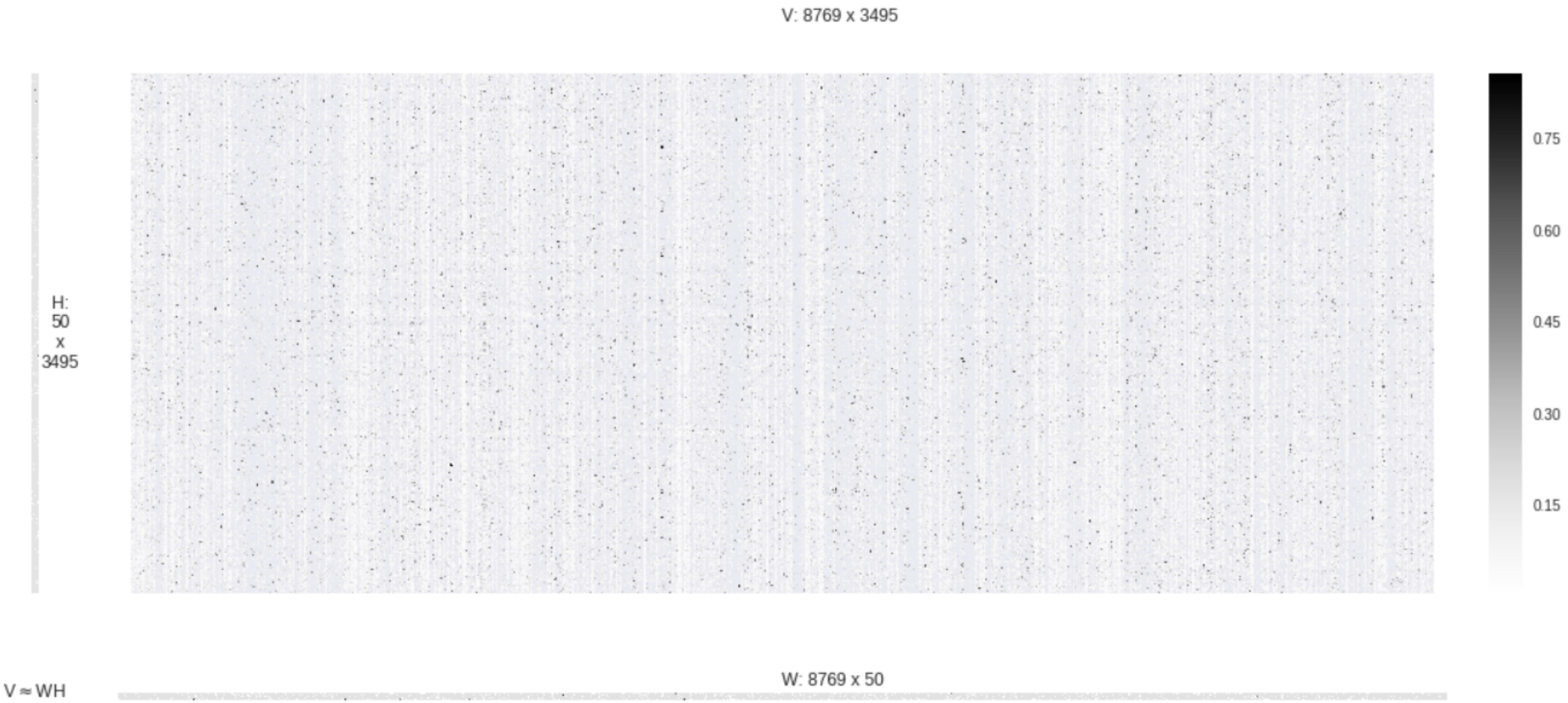


Figure: A topic model of 3495 documents on climate change from the year 2000

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 relationship between the IPCC and scientific literature on climate change? How do the maps relate to the mapped?

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 relationship between the IPCC and scientific literature on climate change? How do the maps relate to the mapped?

Steps

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 relationship between the IPCC and scientific literature on climate change? How do the maps relate to the mapped?

Steps

1. Download document metadata from Web of Science (WoS)

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 relationship between the IPCC and scientific literature on climate change? How do the maps relate to the mapped?

Steps

1. Download document metadata from Web of Science (WoS)
2. Match documents to reference lists from IPCC reports

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 relationship between the IPCC and scientific literature on climate change? How do the maps relate to the mapped?

Steps

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

Data - Query

(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 (Expressio Unius Est Exclusio Alterius)
- WoS not all peer-reviewed literature
- Missing grey literature
- Missing relevant literature not directly about climate change

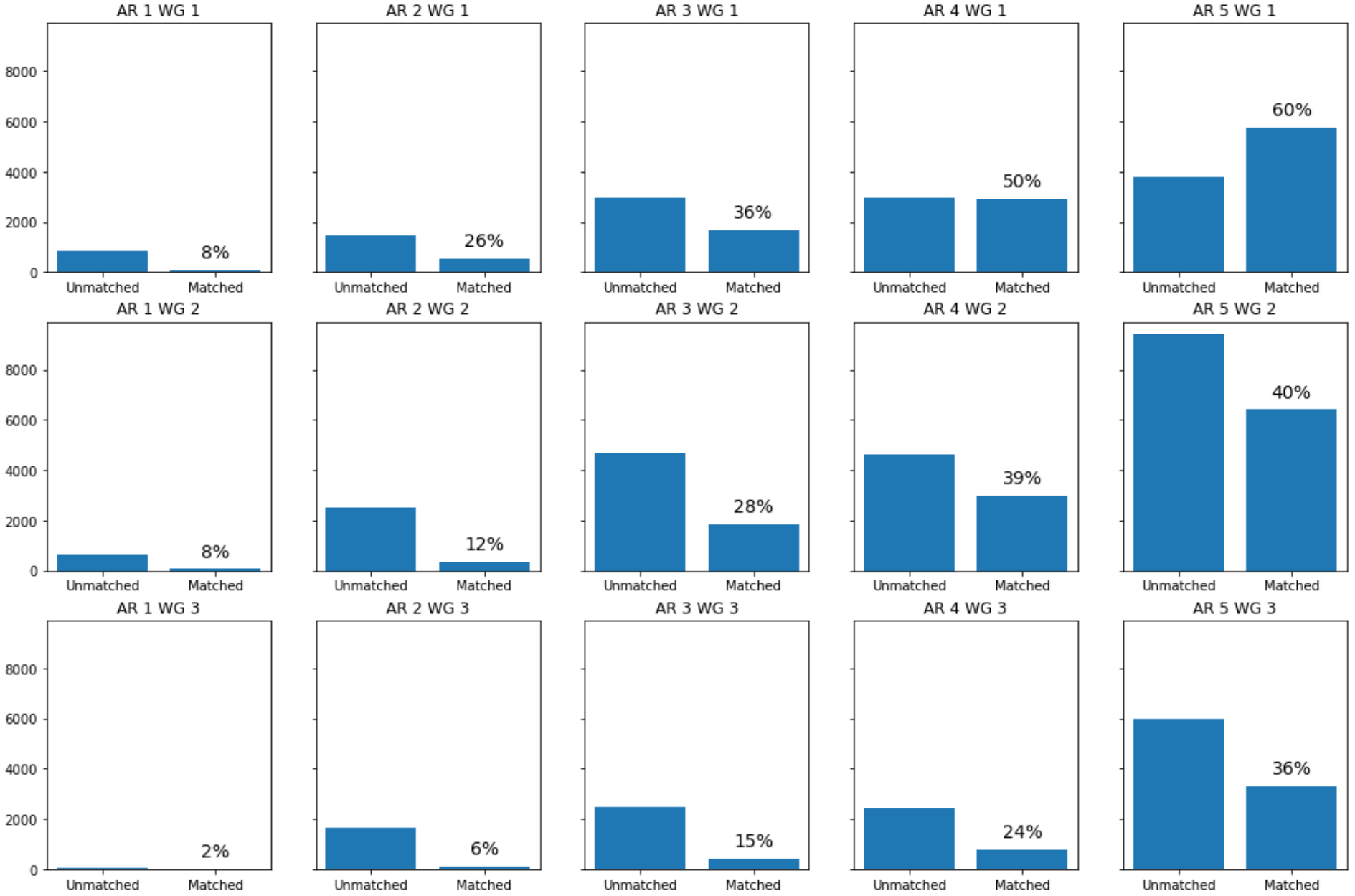
Data - IPCC References

Matching process

For each Reference:

- Check for case-insensitive title matches
- Calculate the Jaccard similarity score for two word shingles 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

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

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 International 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

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

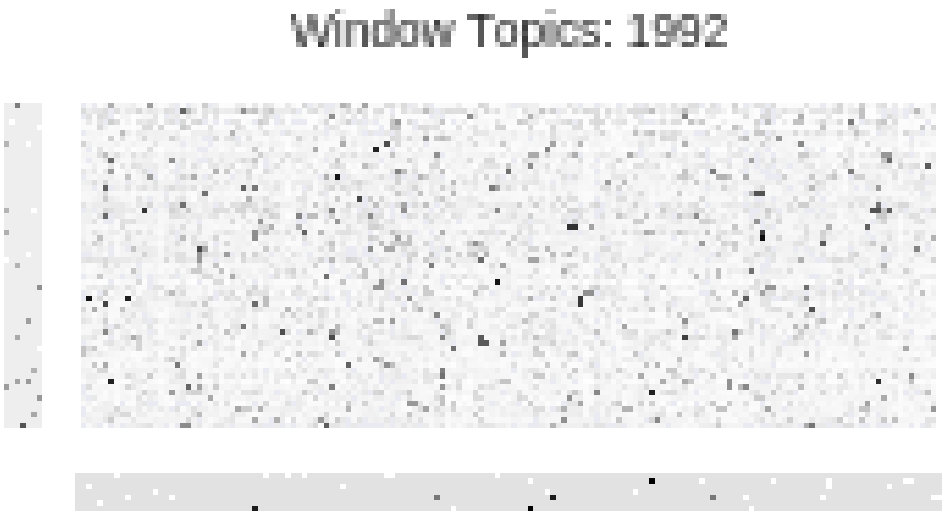
The topic models above assume that the topics, and the words that make them up, are stable over time. Two approaches to better model dynamic topics:

- 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.

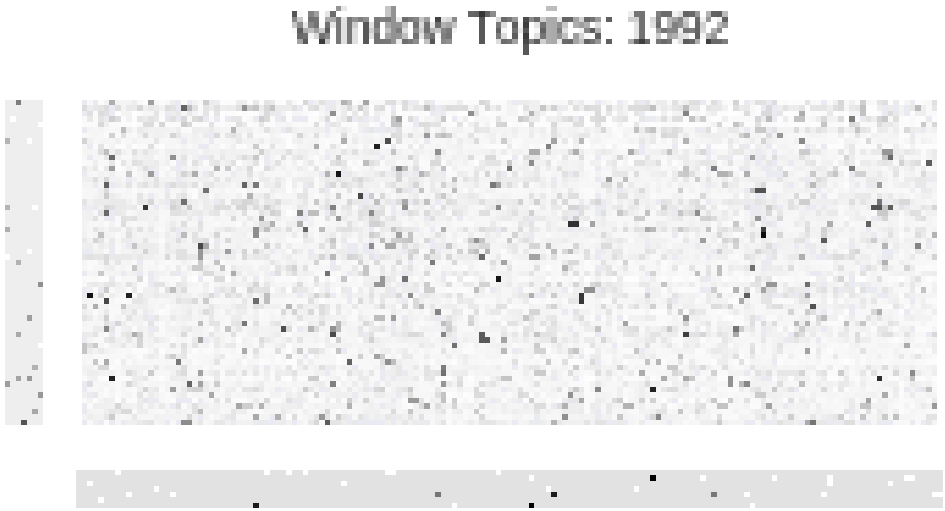
Where the size and variety of the literature we want to model has increased exponentially, we need an approach that allows for the emergence of new topics.

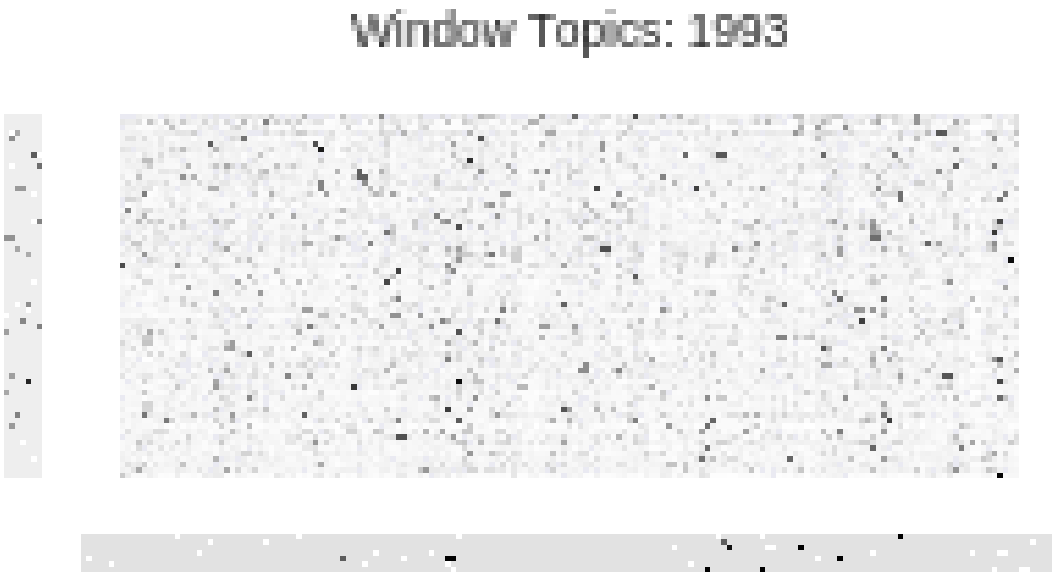
Dynamic NMF

Dynamic NMF



Dynamic NMF

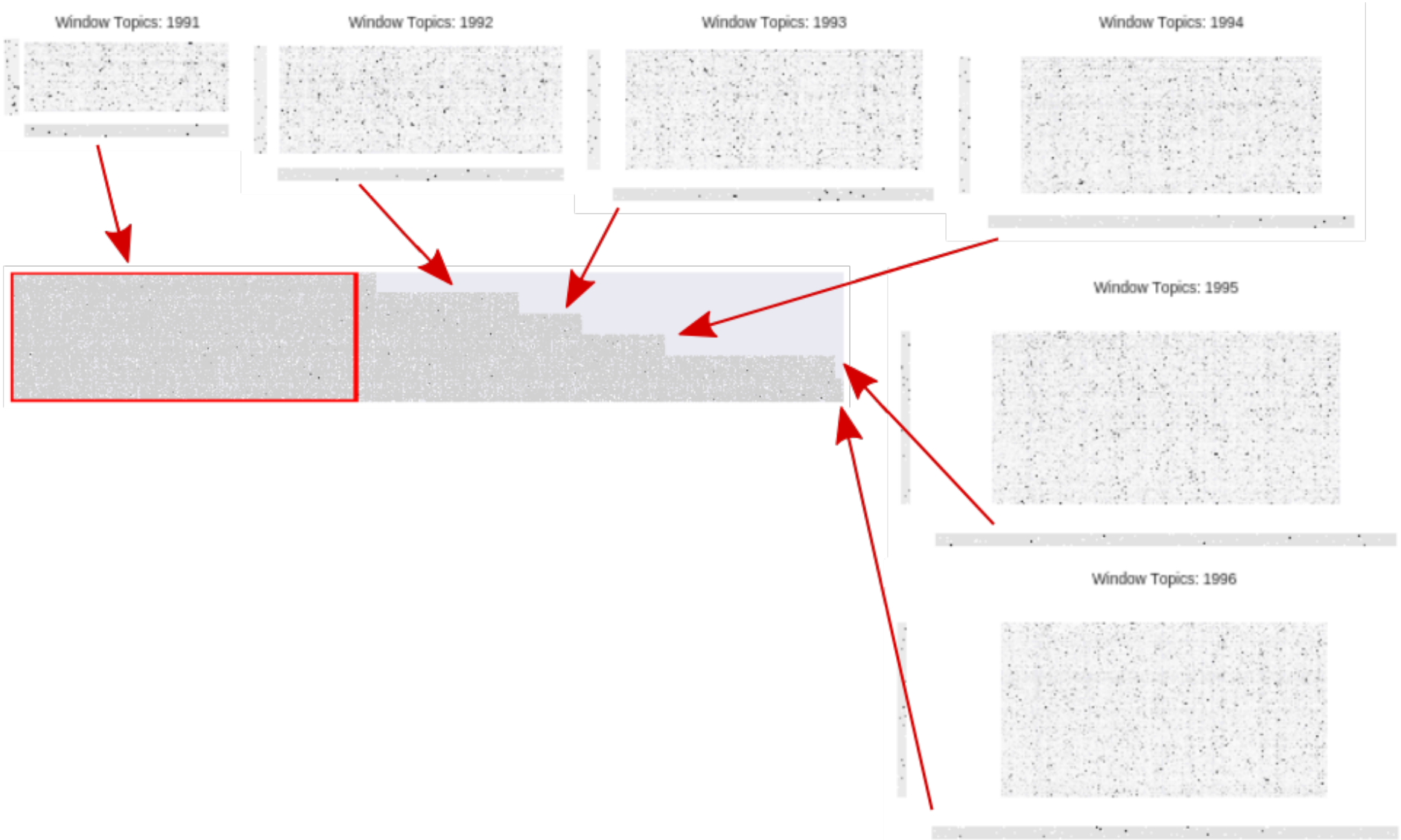




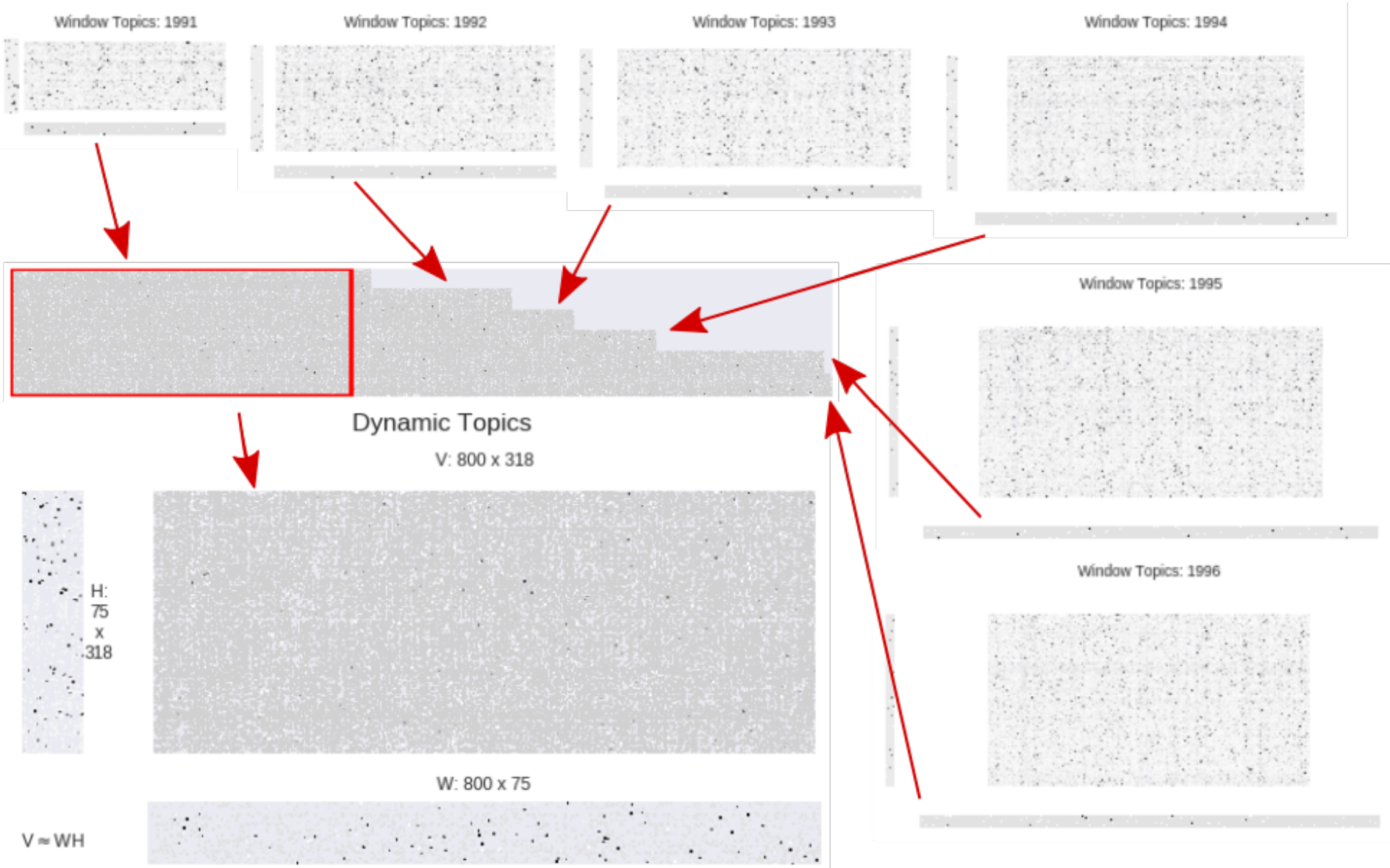
Dynamic NMF

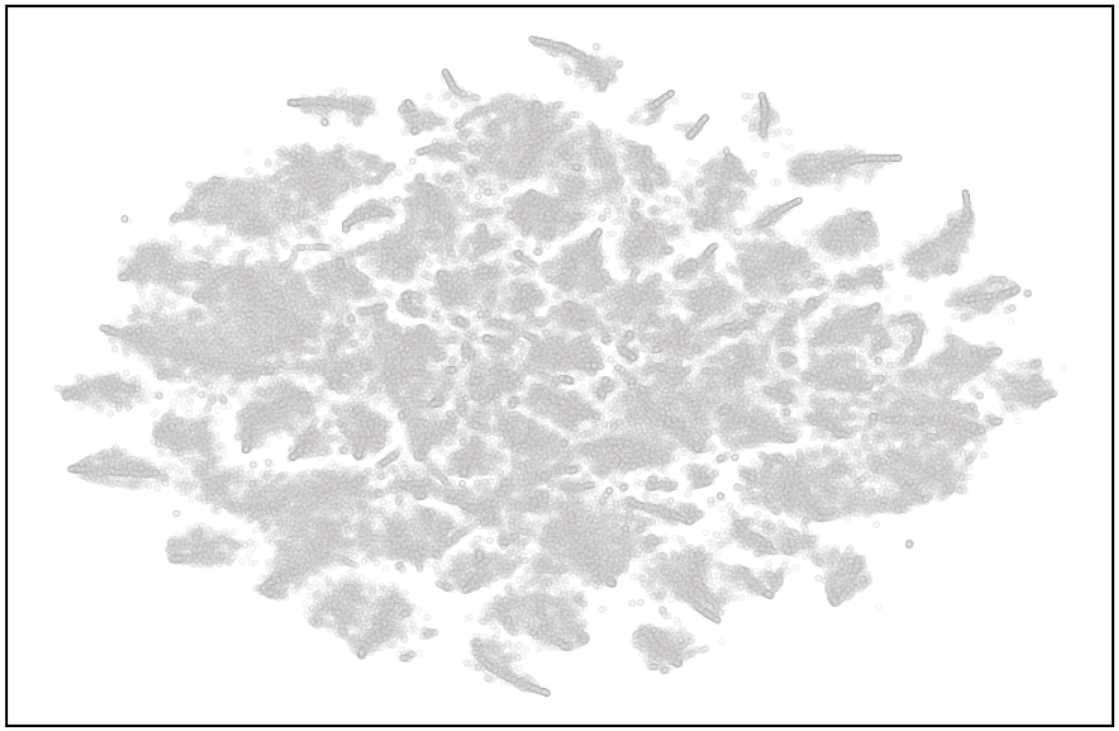

Window Topics: 1991
Window Topics: 1992
Window Topics: 1993
Window Topics: 1994
Window Topics: 1995
Window Topics: 1996

Dynamic NMF


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

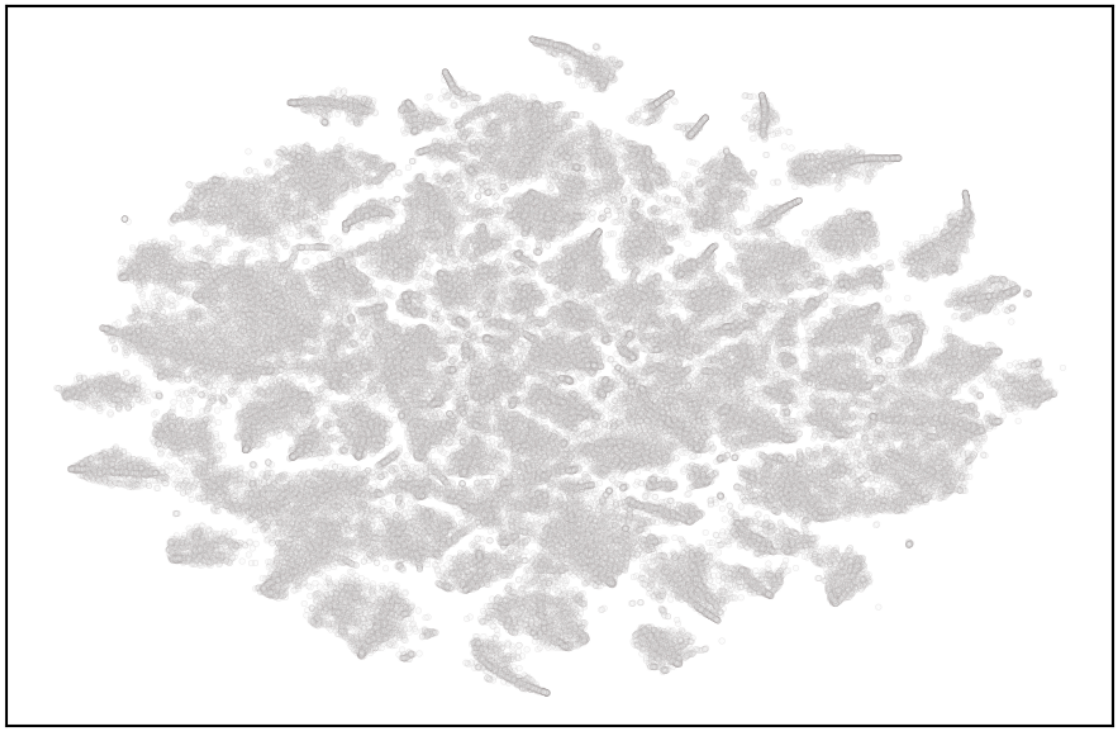
Dynamic NMF


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



Outline

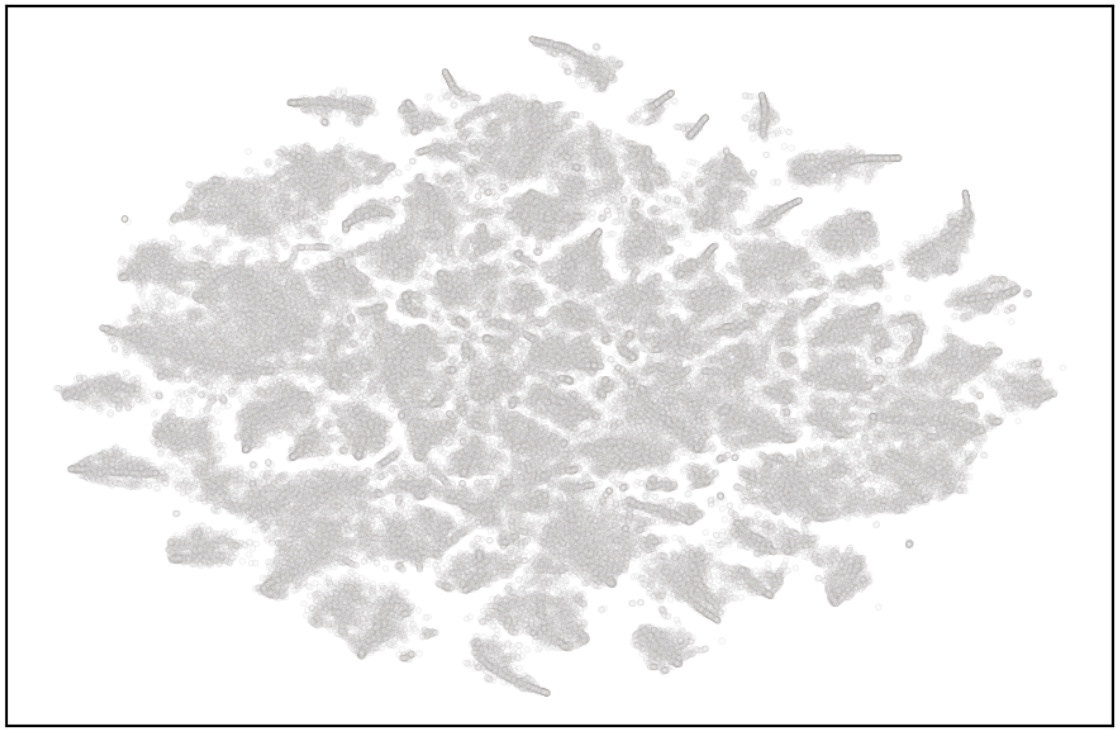
Figure: A topographical representation of 100,000 documents on climate change



Outline

- Topography

Figure: A topographical representation of 100,000 documents on climate change



Outline

- Topography

Figure: A topographical representation of 100,000 documents on climate change

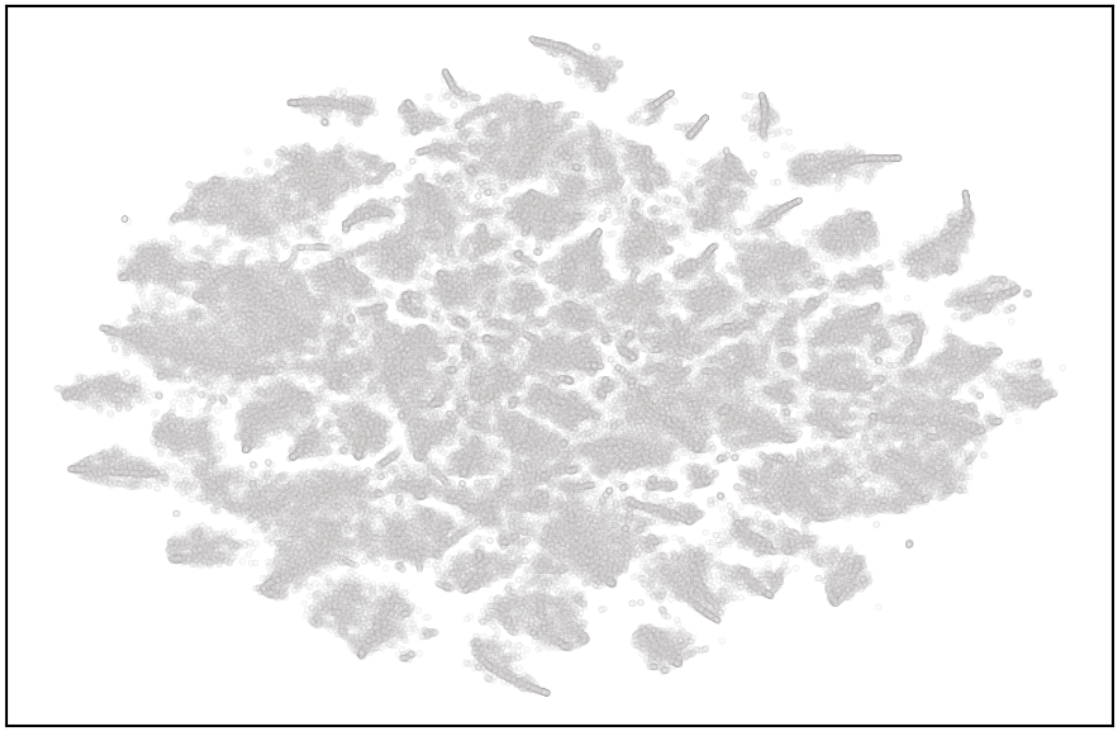
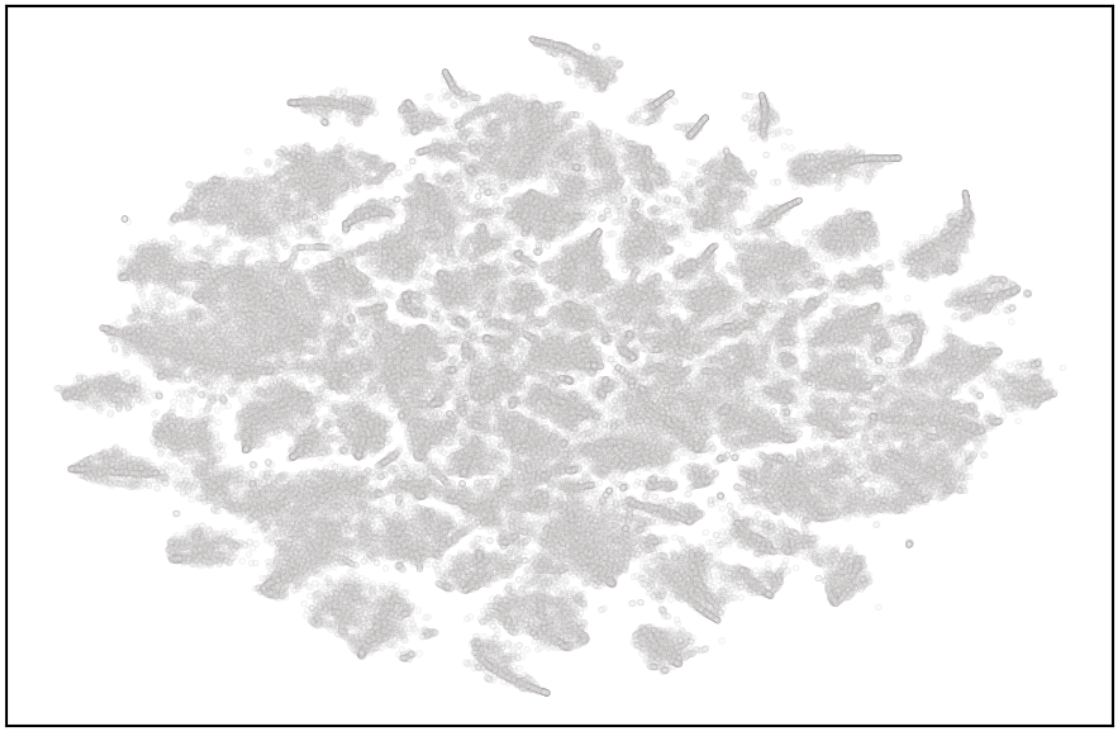


Figure: A topographical representation of 100,000 documents on climate change

Outline

- Topography
- Development



Outline

- Topography
- Development
- Representation in IPCC reports

Figure: A topographical representation of 100,000 documents on climate change

Results - Topography

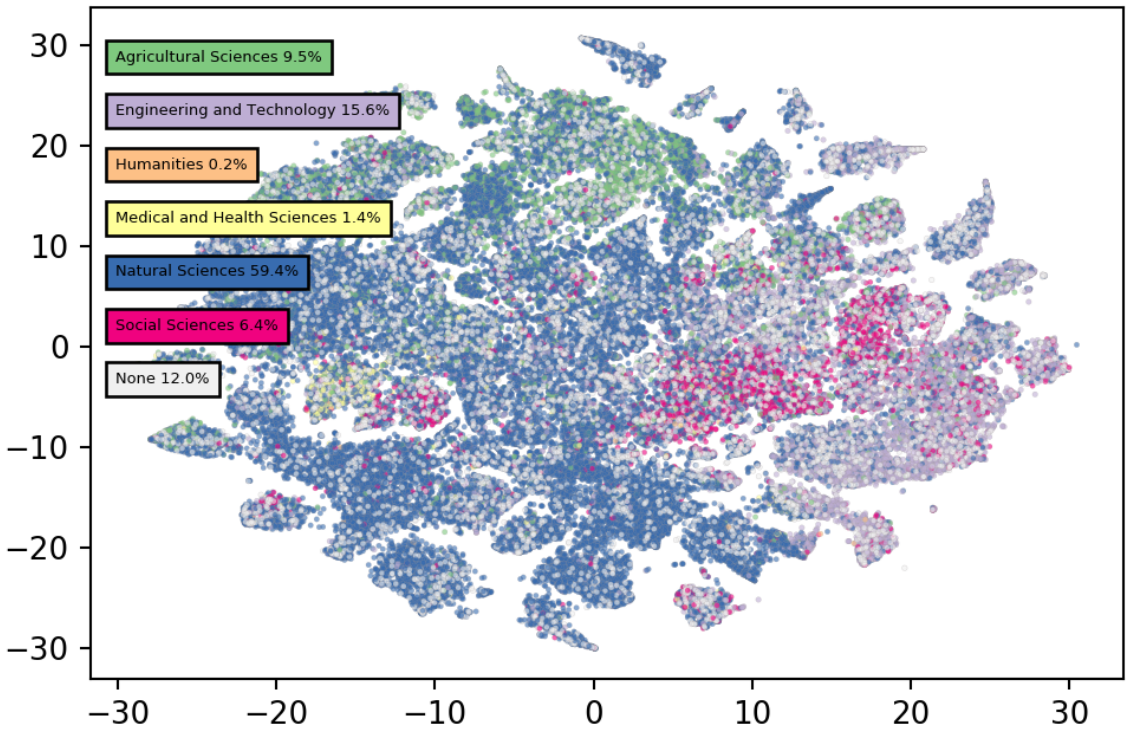


Figure: A topographical representation of 100,000 documents on climate change. Document positions are dictated by topic scores, reduced to two dimensions through t-SNE

Results - Topography

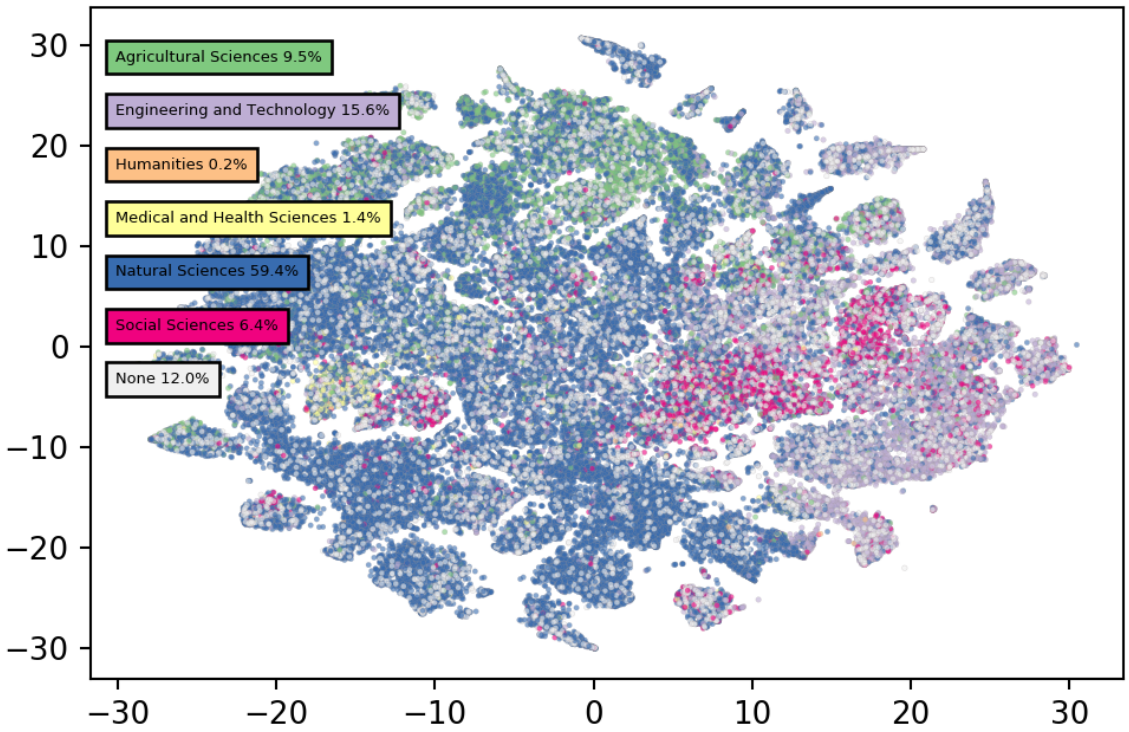


Figure: A topographical representation of 100,000 documents on climate change. Document positions are dictated by topic scores, reduced to two dimensions through t-SNE

- t-distributed stochastic neighbour embedding (t-SNE) (van der Maaten and Hinton, 2008) places nodes with similar topic scores close together

Results - Topography

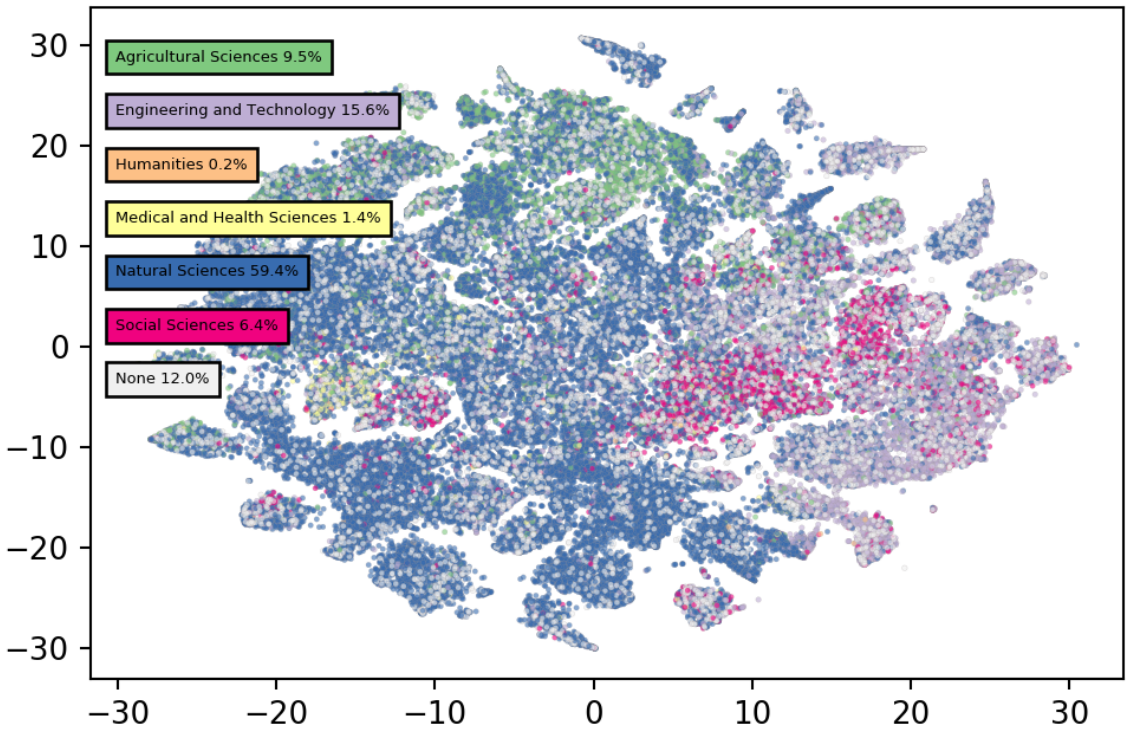


Figure: A topographical representation of 100,000 documents on climate change. Document positions are dictated by topic scores, reduced to two dimensions through t-SNE

- t-distributed stochastic neighbour embedding (t-SNE) (van der Maaten and Hinton, 2008) places nodes with similar topic scores close together
- This *projects* the 100-dimensional topic scores onto a conventional 2-dimensional topographical map

Results - Topography

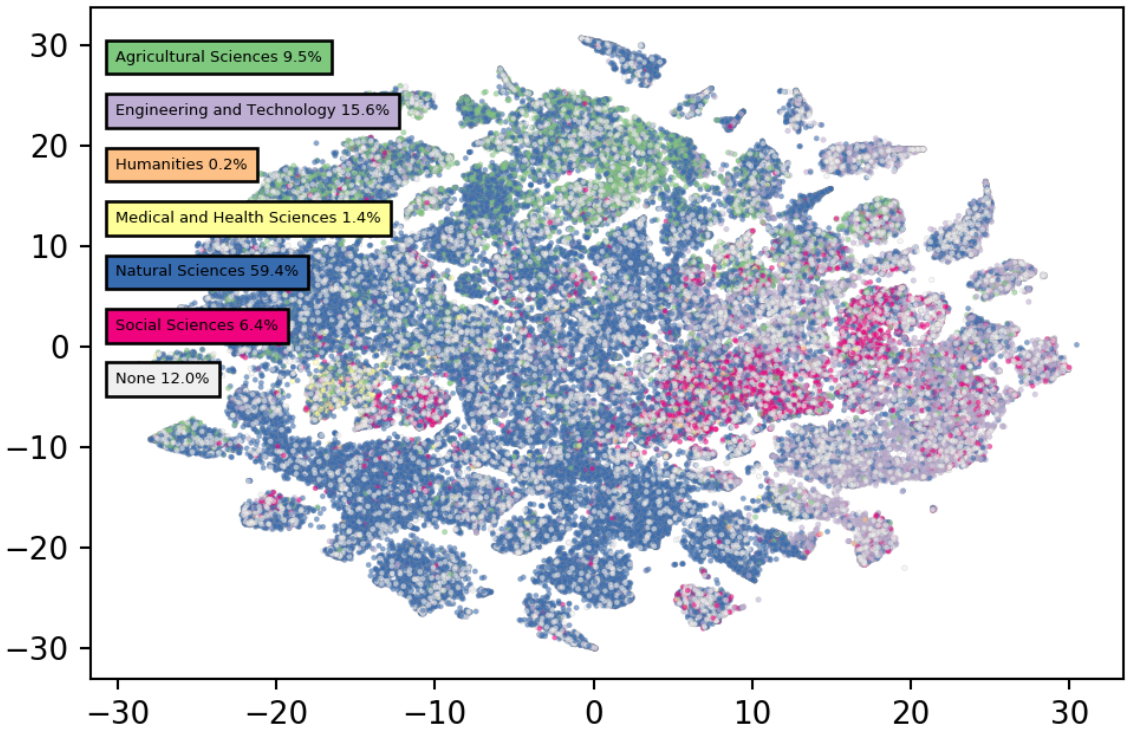
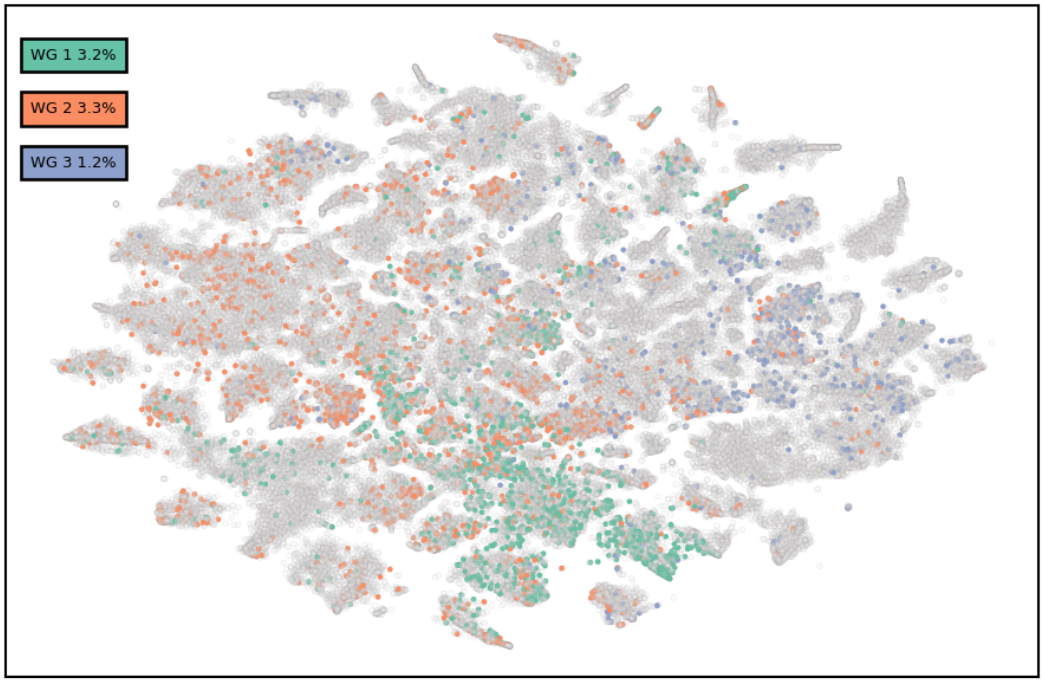


Figure: A topographical representation of 100,000 documents on climate change. Document positions are dictated by topic scores, reduced to two dimensions through t-SNE

- t-distributed stochastic neighbour embedding (t-SNE) (van der Maaten and Hinton, 2008) places nodes with similar topic scores close together
- This *projects* the 100-dimensional topic scores onto a conventional 2-dimensional topographical map
- Topics cut across journal disciplinary categories, but the disciplinary structure remains visible

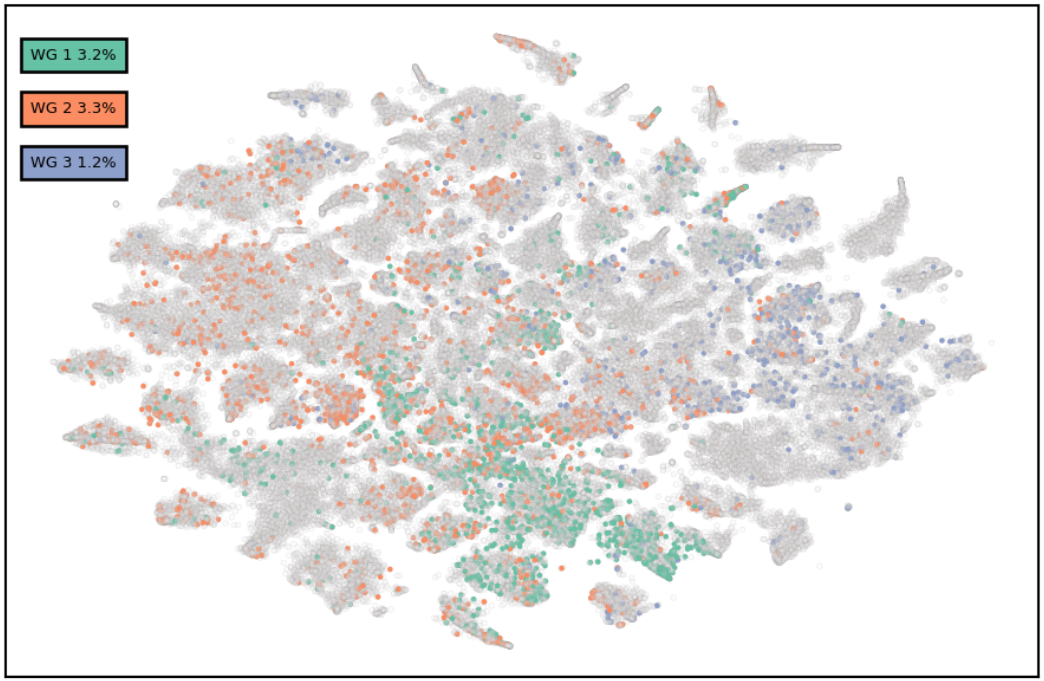
Results - Topography



- The disciplinary structure fits with the pattern of citations by IPCC reports

Figure: A topographical representation of 100,000 documents on climate change. Document positions are dictated by topic scores, reduced to two dimensions through t-SNE, nodes are coloured by references by IPCC reports

Results - Topography



- The disciplinary structure fits with the pattern of citations by IPCC reports
- Citations by IPCC reports are densest in WGI/Natural science topical areas.

Figure: A topographical representation of 100,000 documents on climate change. Document positions are dictated by topic scores, reduced to two dimensions through t-SNE, nodes are coloured by references by IPCC reports

Results - IPCC Representation

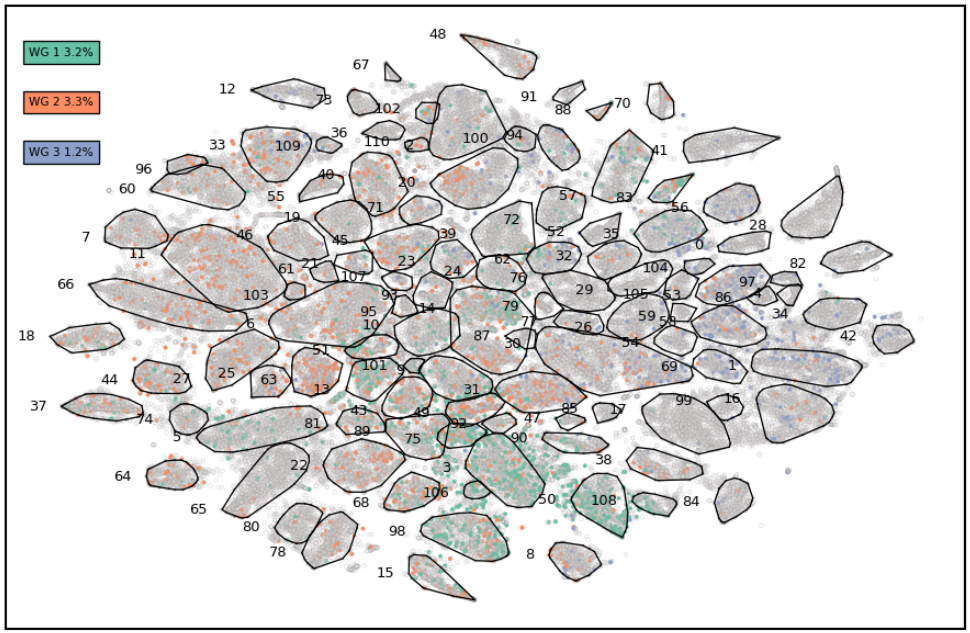


Figure: Clusters of documents

- We can identify clusters of documents, each of which are dominated by topics or clusters of topics

Results - IPCC Representation

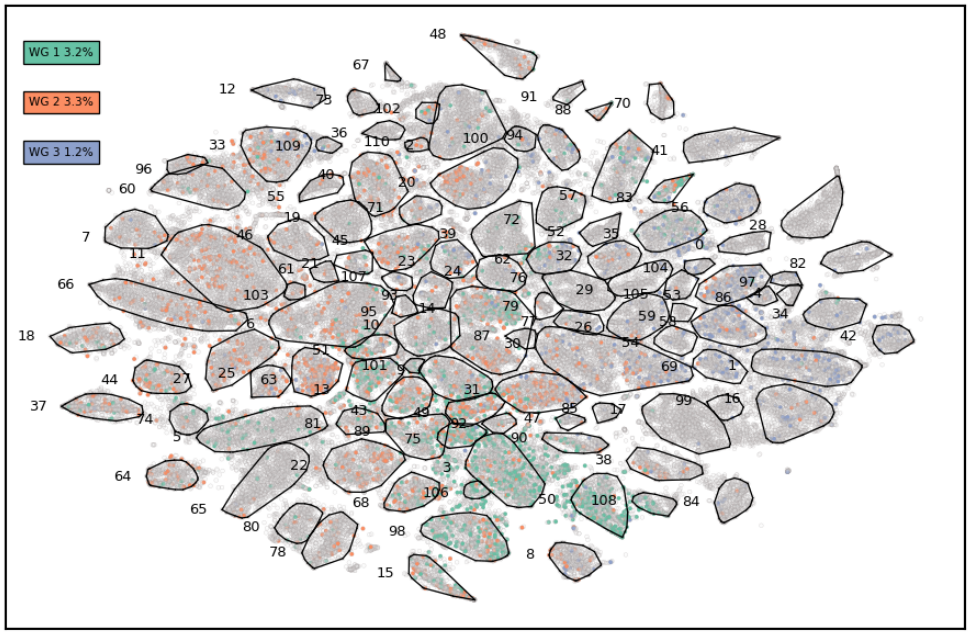


Figure: Clusters of documents

- We can identify clusters of documents, each of which are dominated by topics or clusters of topics
- 58 is a mixture of economics, policies and taxes, and emissions reductions

Results - IPCC Representation

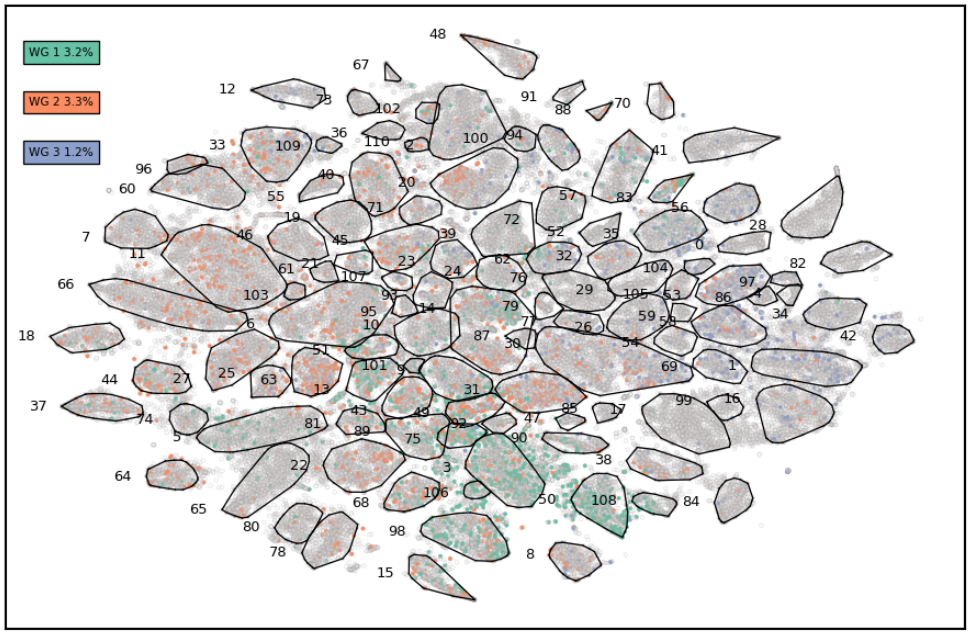


Figure: Clusters of documents

- We can identify clusters of documents, each of which are dominated by topics or clusters of topics
- 58 is a mixture of economics, policies and taxes, and emissions reductions
- 41 is on CO2 storage

Results - IPCC Representation

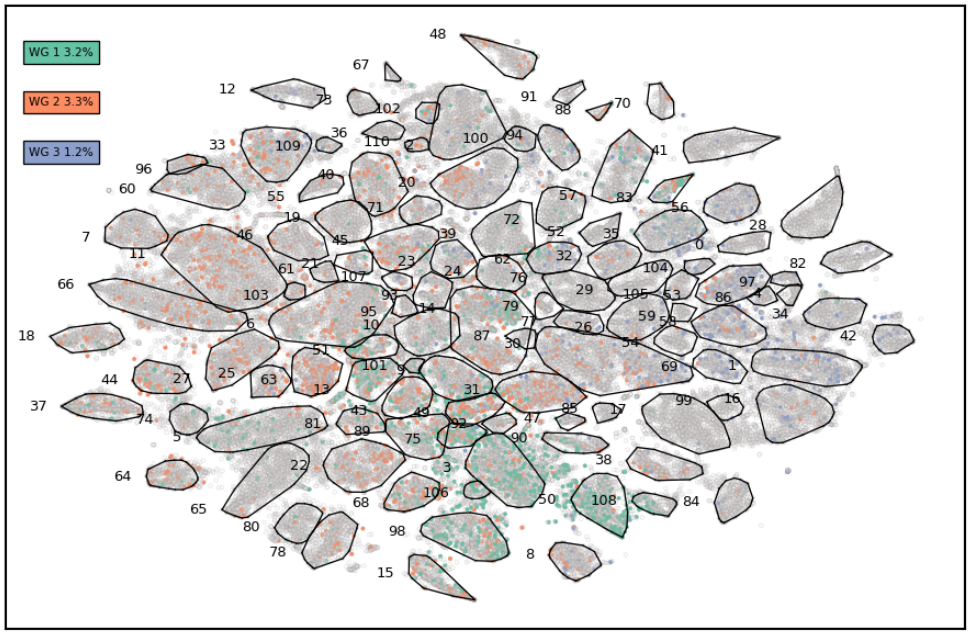


Figure: Clusters of documents

- We can identify clusters of documents, each of which are dominated by topics or clusters of topics
- 58 is a mixture of economics, policies and taxes, and emissions reductions
- 41 is on CO2 storage
- 11 is on species and habitats

Results - IPCC Representation

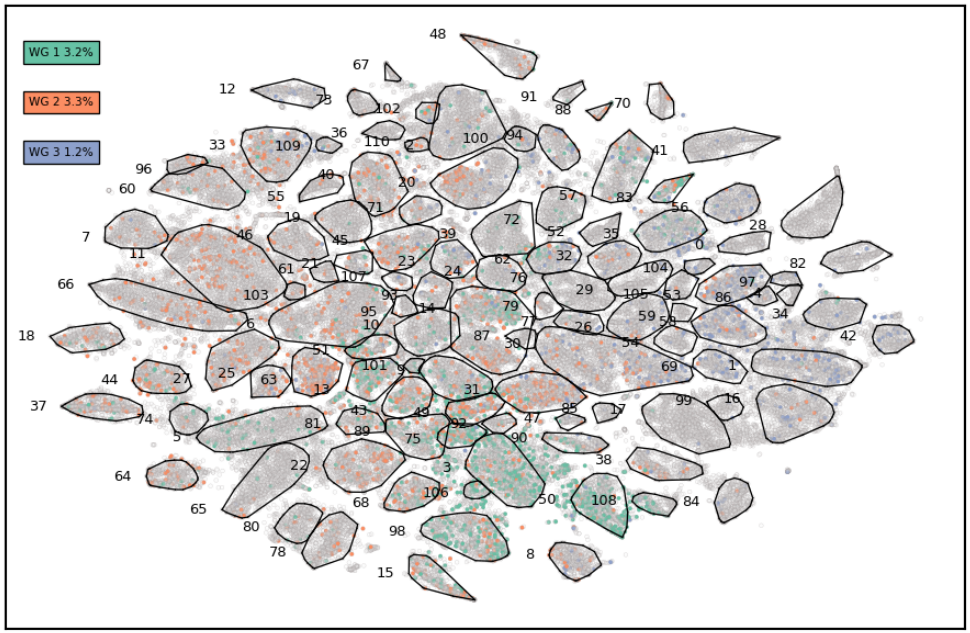
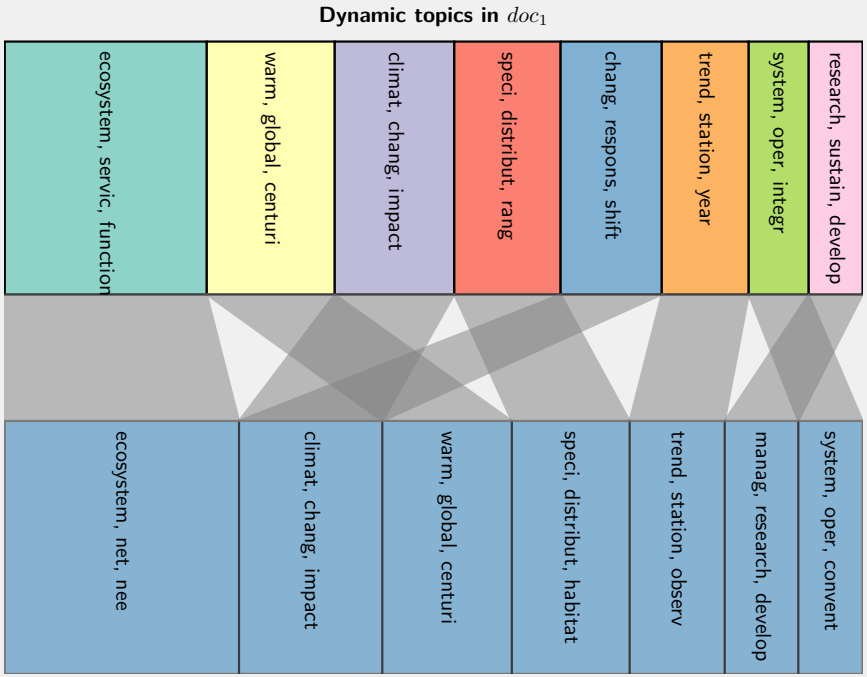
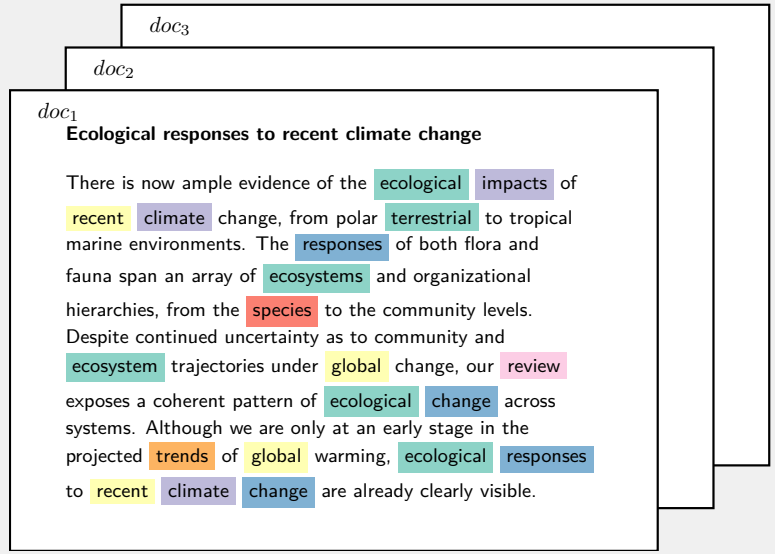


Figure: Clusters of documents

- We can identify clusters of documents, each of which are dominated by topics or clusters of topics
- 58 is a mixture of economics, policies and taxes, and emissions reductions
- 41 is on CO2 storage
- 11 is on species and habitats
- 33 is on forests

Results - Topography





Topic Doc	ecosystem net nee	climat chang impact	warm global centuri	speci distribut habitat
doc_1	0.034	0.021	0.019	0.017
doc_2	...	...	...	...
doc_3	...	...	...	...

Doc Topic Matrix

×

Term Topic	chang	ecolog	global	recent
ecosystem net nee	0.18	0.39	0.01	0.03
climat chang impact	6.33	0	0.22	0.19
warm global centuri	0	0	3.72	0.31

Topic Term Matrix

≈

Term Doc	chang	ecolog	global	recent
doc_1	4	3	2	2
doc_2	...	...	...	...
doc_3	...	...	...	...
doc_4	...	...	...	...

Doc Term Matrix

Figure: How topics describe documents

Results - Development

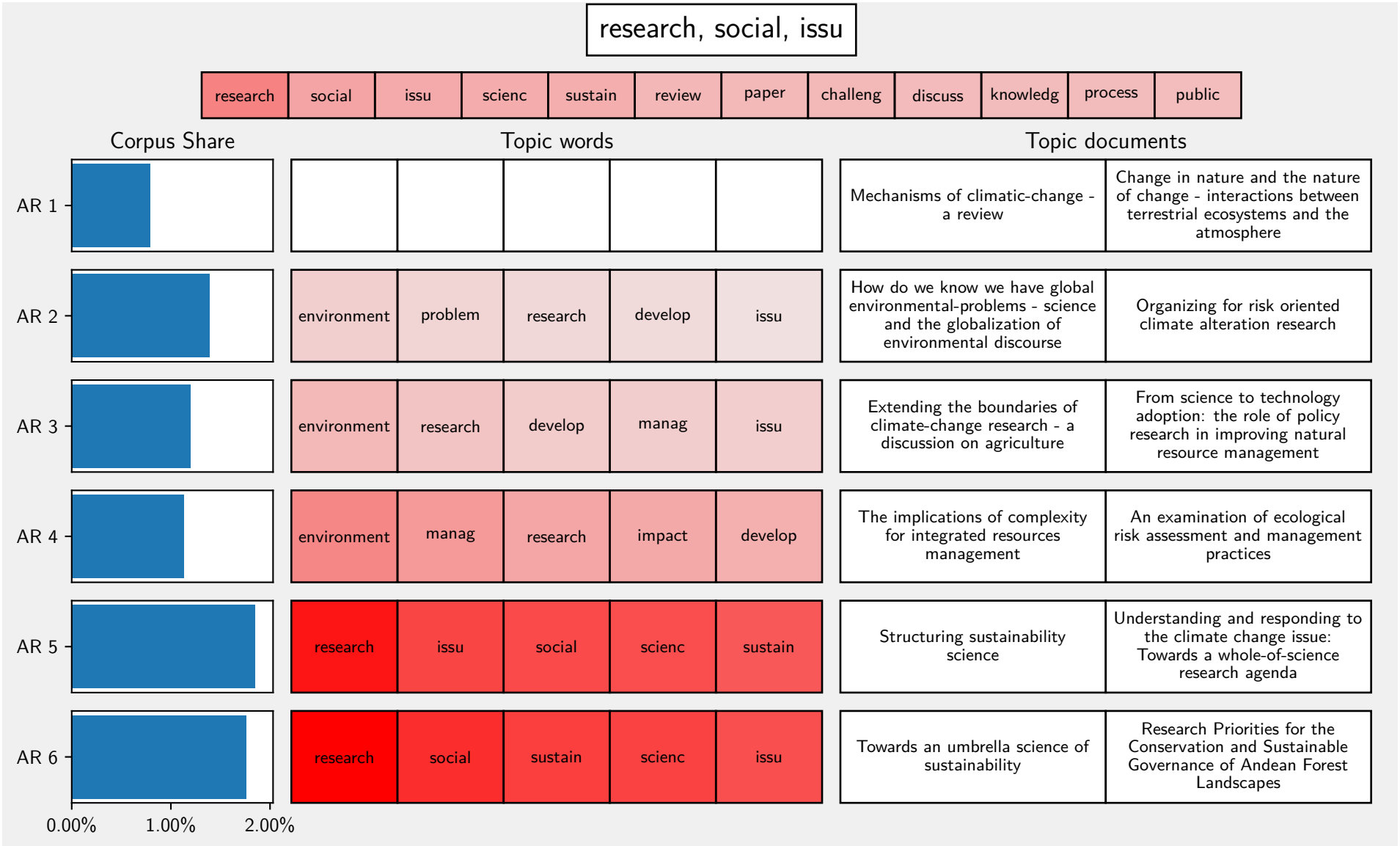


Figure: The evolution of topics

Results - Development

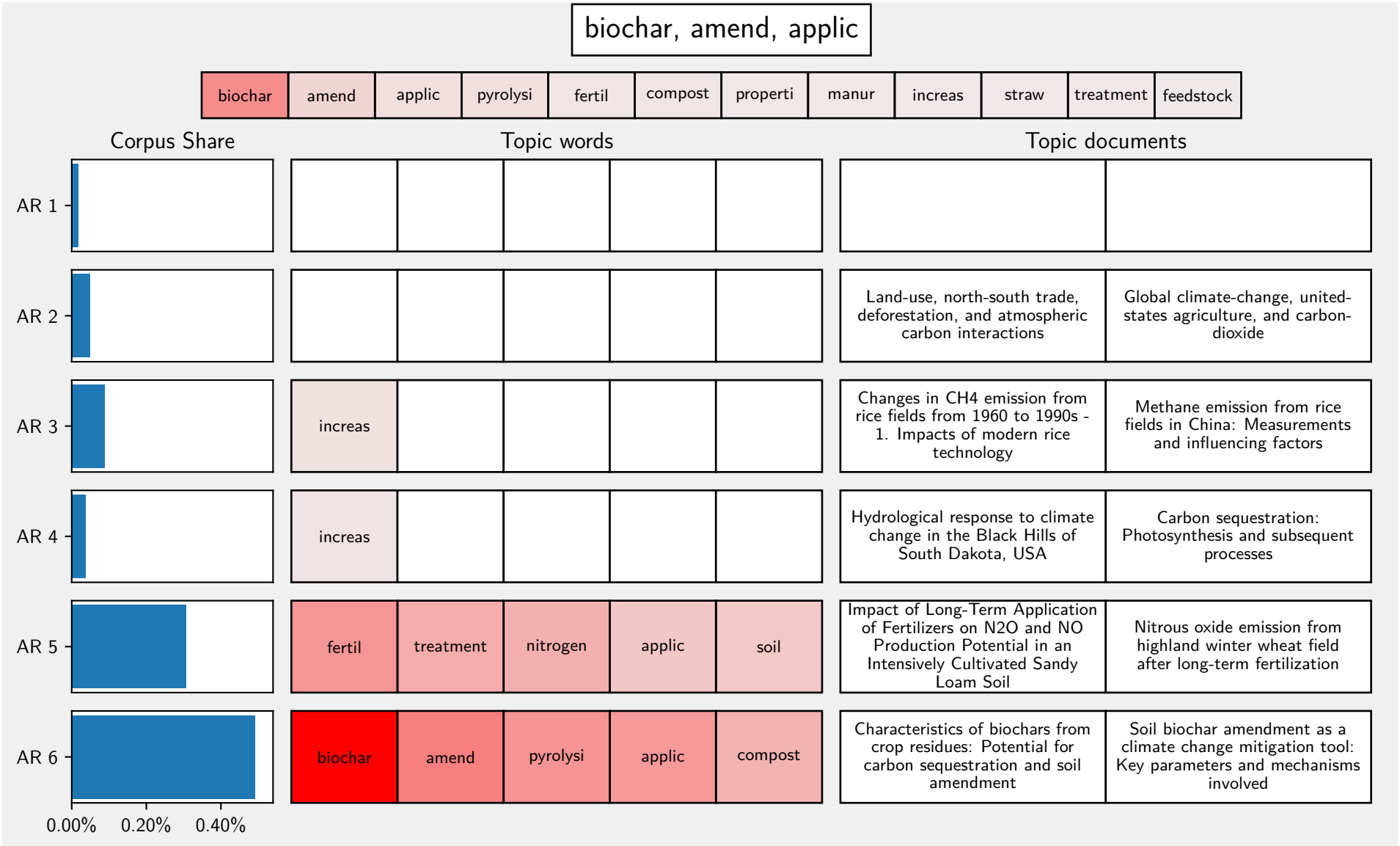
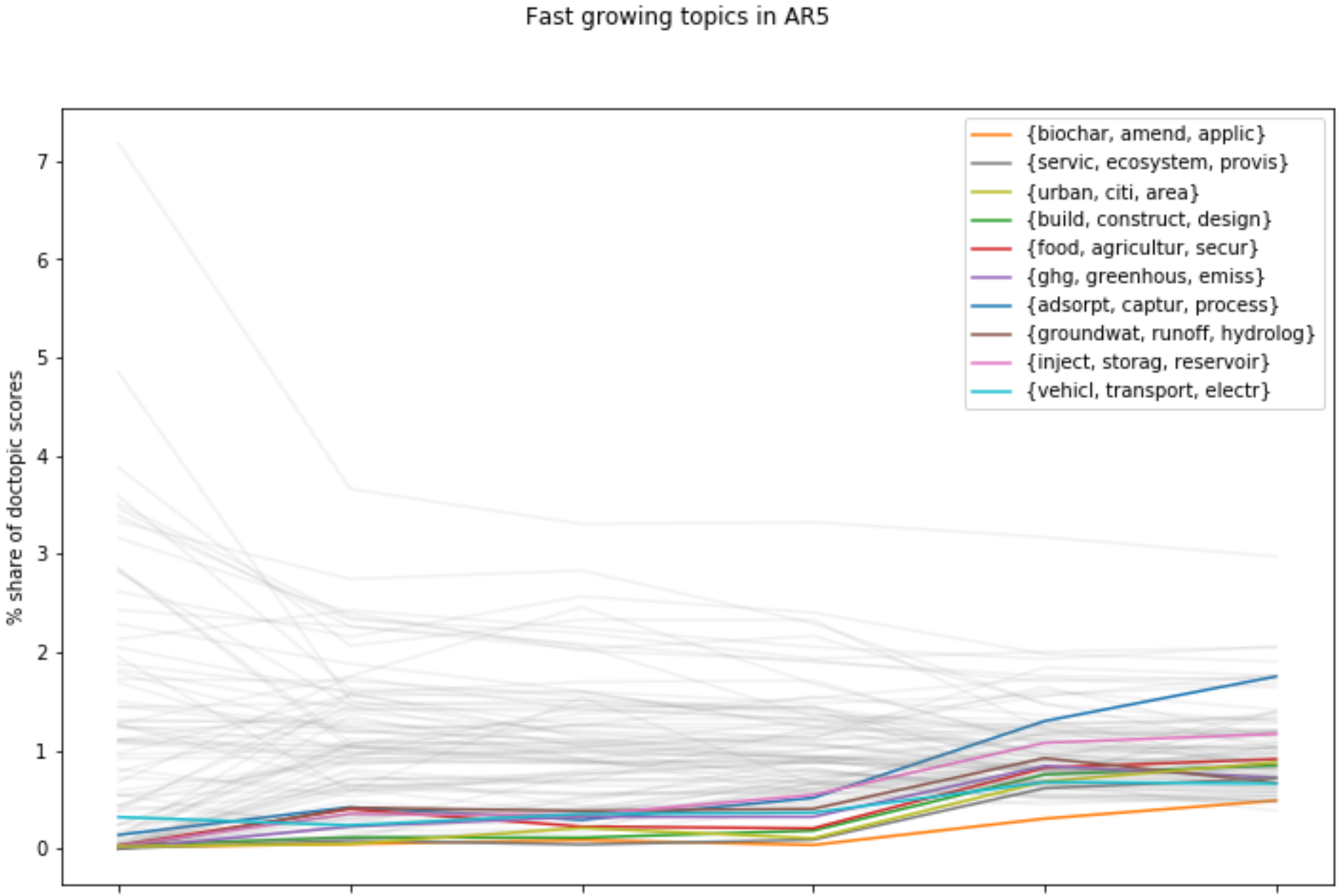


Figure: The emergence of new topics

Results - Development



Results - IPCC representation

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

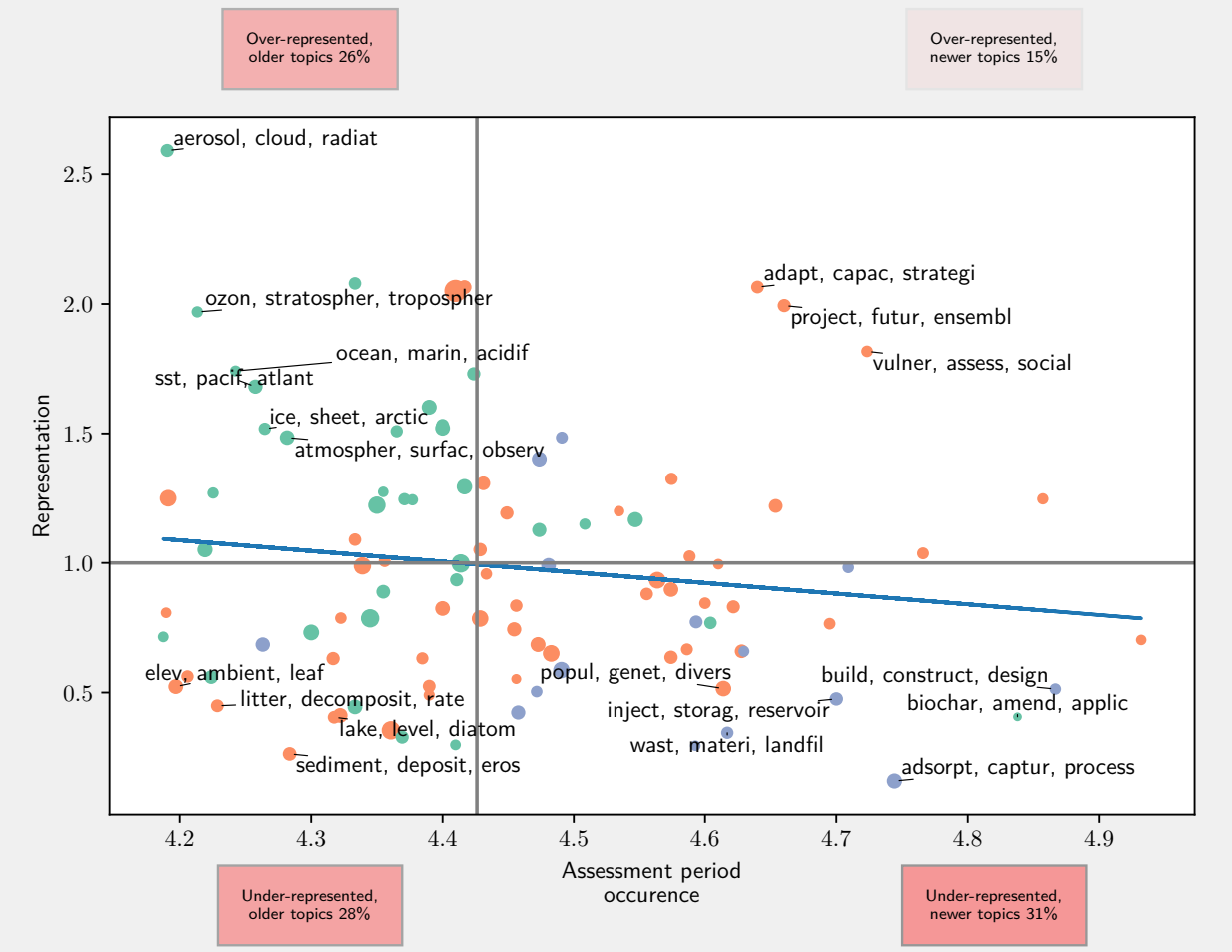
Results - IPCC representation

- 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 each topic's share in the IPCC sample by its share in the whole corpus

Results - IPCC representation

- 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 each topic's share in the IPCC sample by its share in the whole corpus
- This and other measures come from literature on the proportionality of electoral systems (e.g Karpov, 2008)

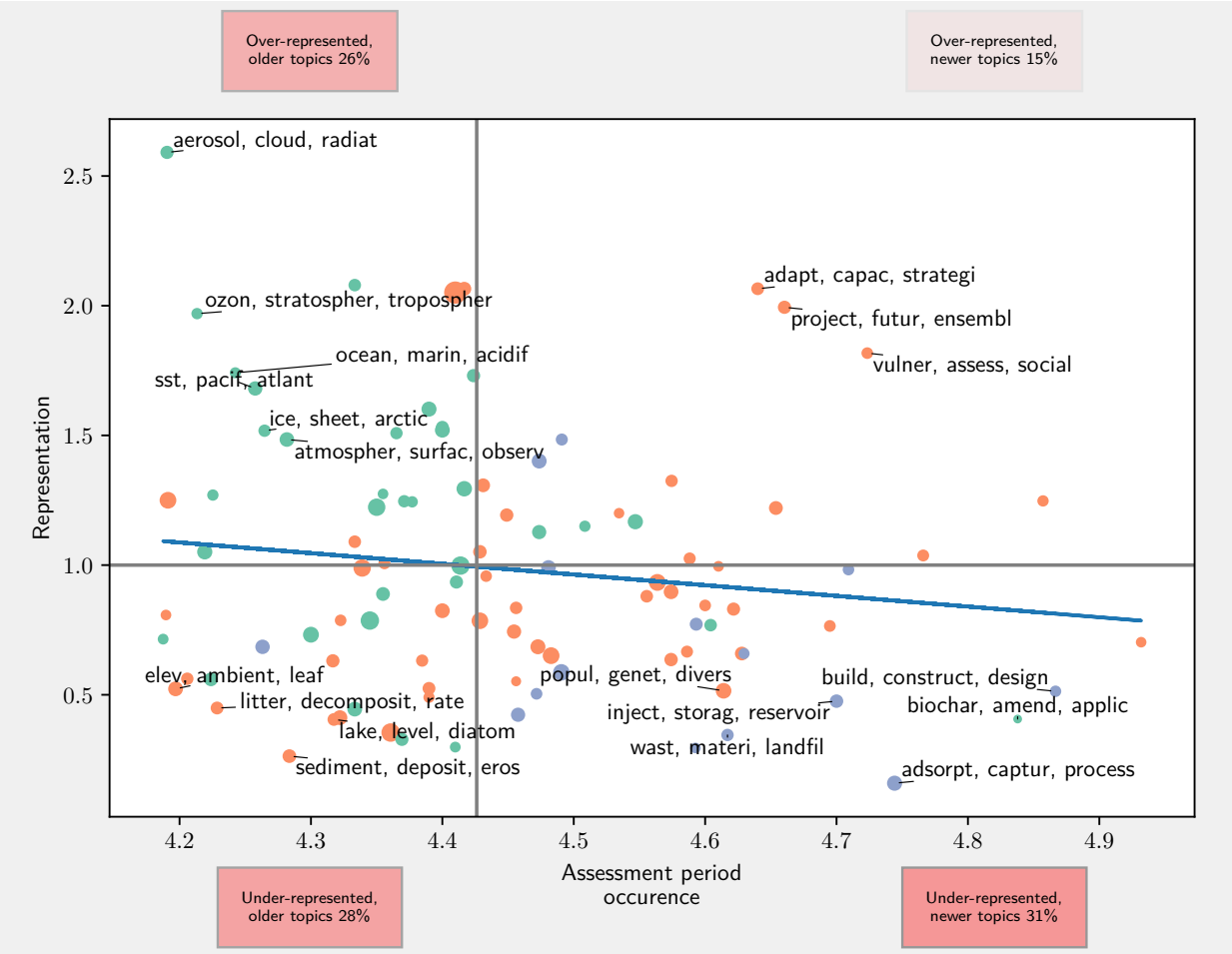
Results - IPCC Representation



- Topics on the scientific basis of climate change are well represented in IPCC reports and have been around longer than average

Figure: New and old, better and less well represented topics

Results - IPCC Representation



- Topics on the scientific basis of climate change are well represented in IPCC reports and have been around longer than average
- Topics on impacts have seen lots of recent growth, and are well represented in IPCC reports

Figure: New and old, better and less well represented topics

Results - IPCC Representation

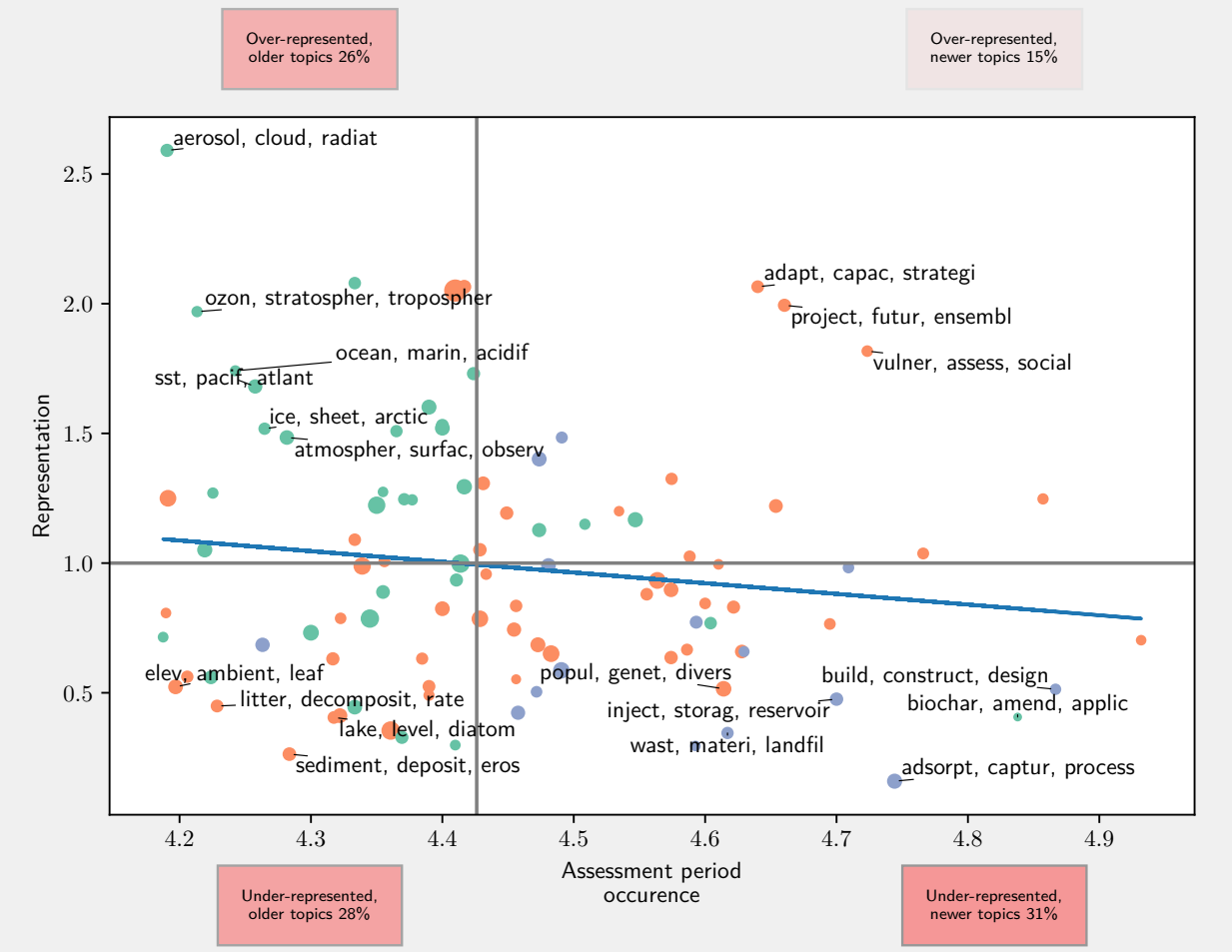


Figure: New and old, better and less well represented topics

- Topics on the scientific basis of climate change are well represented in IPCC reports and have been around longer than average
- Topics on impacts have seen lots of recent growth, and are well represented in IPCC reports
- Topics on "solutions", or on new technologies or contexts for mitigation have seen lots of recent growth, but are under-represented in IPCC reports

Robustness

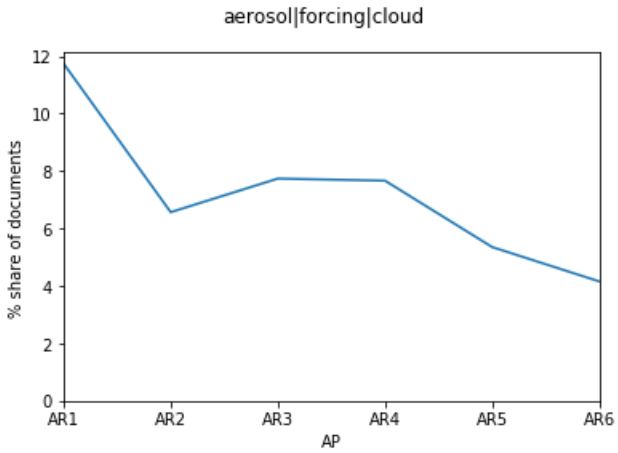
Most work focuses on the measuring the sensitivity of the reconstruction error to different framings and parameters of the model. What we are more interested in is checking if the key messages are sound.

Robustness

Most work focuses on the measuring the sensitivity of the reconstruction error to different framings and parameters of the model. What we are more interested in is checking if the key messages are sound.

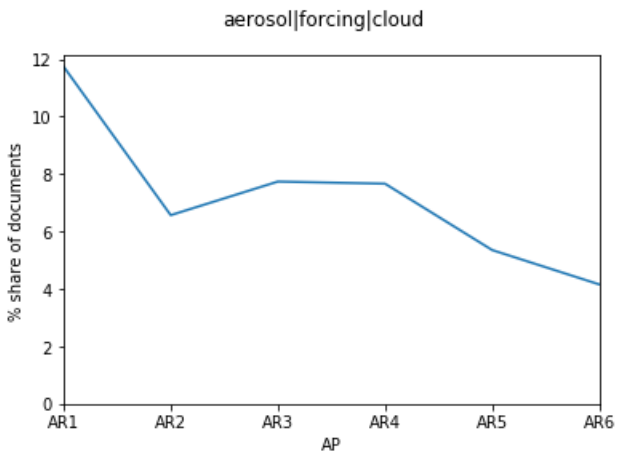
Robustness

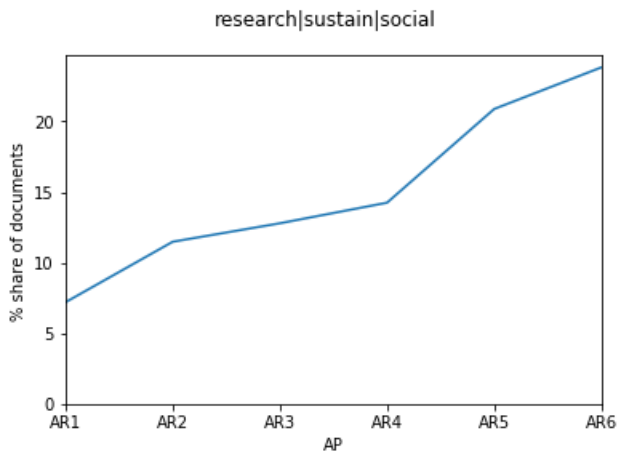
Most work focuses on the measuring the sensitivity of the reconstruction error to different framings and parameters of the model. What we are more interested in is checking if the key messages are sound.



Robustness

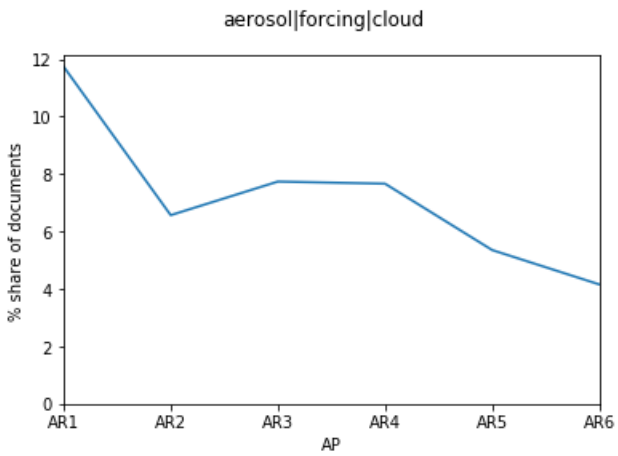
Most work focuses on the measuring the sensitivity of the reconstruction error to different framings and parameters of the model. What we are more interested in is checking if the key messages are sound.

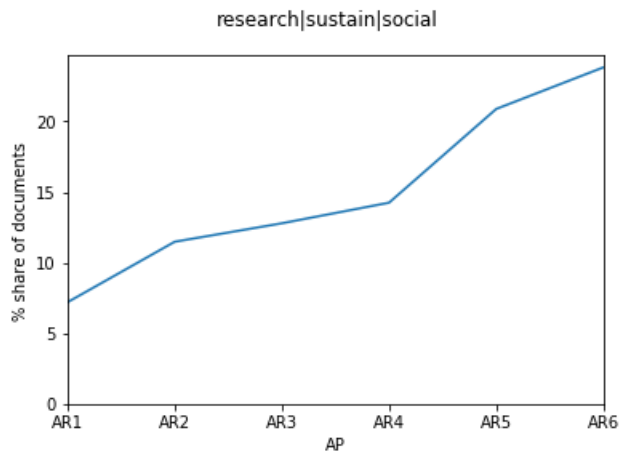


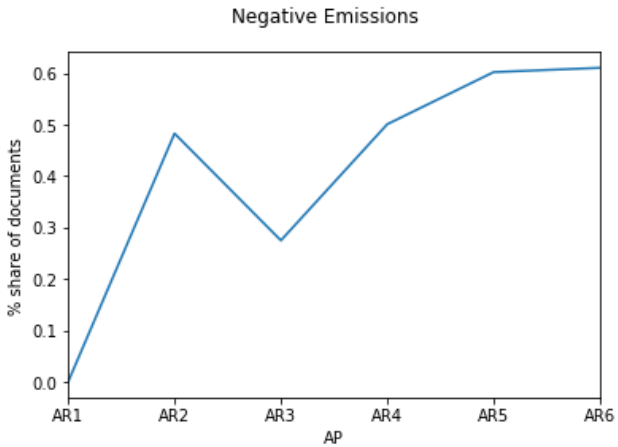


Robustness

Most work focuses on the measuring the sensitivity of the reconstruction error to different framings and parameters of the model. What we are more interested in is checking if the key messages are sound.

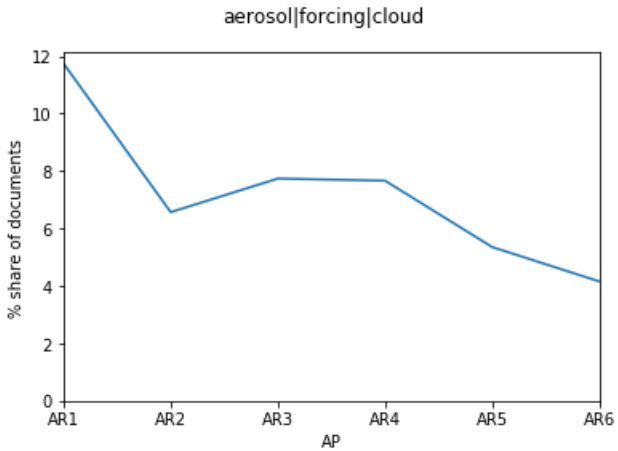


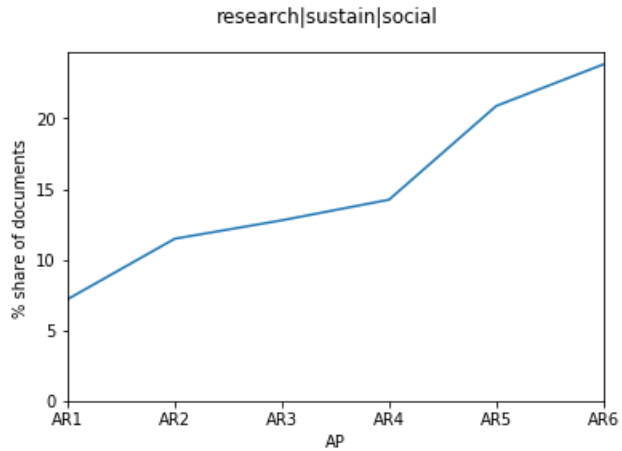


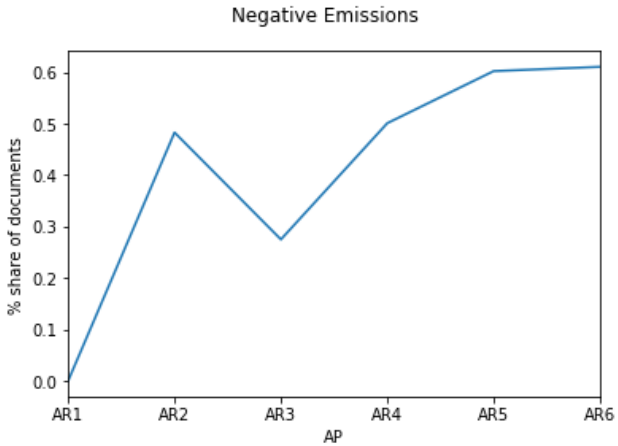


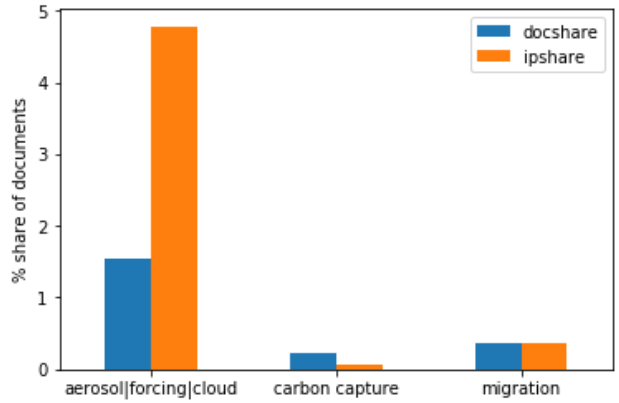
Robustness

Most work focuses on the measuring the sensitivity of the reconstruction error to different framings and parameters of the model. What we are more interested in is checking if the key messages are sound.









Conclusions

- Endogenously discovered topics map and make substantive sense of an unmanageable dataset of climate-relevant literature

Conclusions

- Endogenously discovered topics map and 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 specific, fast growing topics such those on negative emissions

Conclusions

- Endogenously discovered topics map and 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 specific, fast growing topics such those on negative emissions
- Some quantitative evidence is found to support policy makers' dissatisfaction with a lack of 'solution orientation' in IPCC reports (Kowarsch et al., 2017)

Discussion

- A perfectly representative relationship between the literature and IPCC reports is not necessarily desirable

Discussion

- A perfectly representative relationship between the literature and IPCC reports is not necessarily desirable
- The IPCC are best placed to decide which aspects of the literature to emphasise

Discussion

- A perfectly representative relationship between the literature and IPCC reports is not necessarily desirable
- The IPCC are best placed to decide which aspects of the literature to emphasise
- A topography of the literature helps to address issues of emphasis from a point of understanding, and to make decisions clear and transparent

Discussion

- A perfectly representative relationship between the literature and IPCC reports is not necessarily desirable
- The IPCC are best placed to decide which aspects of the literature to emphasise
- A topography of the literature helps to address issues of emphasis from a point of understanding, and to make decisions clear and transparent
- More generally, maps like this present exciting opportunities to aid the process of literature selection, and to understand the science policy process

A Topography of Climate Change Research

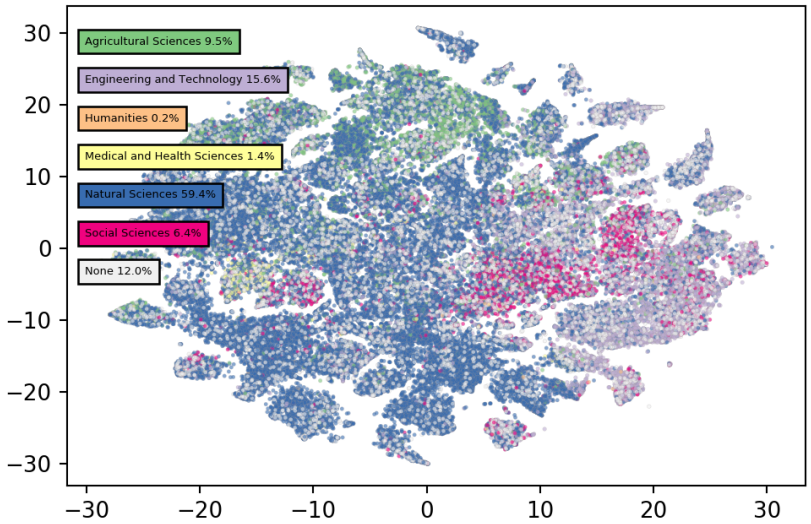


Figure: Portrait of map-makers, Gerard Mercator and Jodocus Hondius (Jodocus Hondius) source: `https://commons.wikimedia.org/wiki/File:Hondius_Portrait_of_map-makers.jpg`

Bibliography

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.

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

Karpov, A. (2008). Measurement of disproportionality in proportional representation systems. *Mathematical and Computer Modelling*, 48(9-10):1421–1438.

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*.

van der Maaten, L. and Hinton, G. (2008). Visualizing Data using t-SNE Laurens. *Journal of Machine Learning Research*, 9:2579–2605.

Robustness

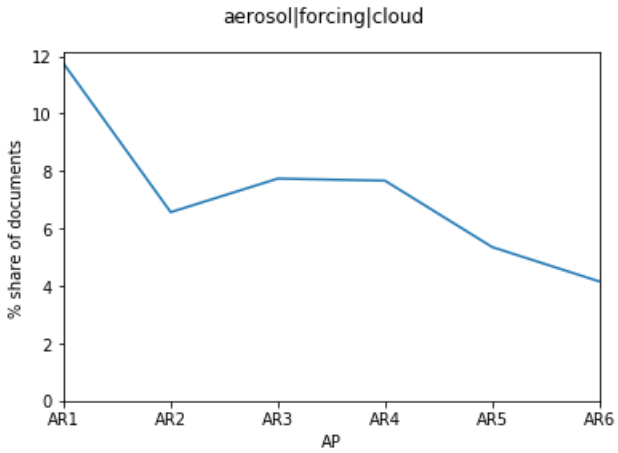
Most work focuses on the measuring the sensitivity of the reconstruction error to different framings and parameters of the model. What we are more interested in is checking if the key messages are sound.

Robustness

Most work focuses on the measuring the sensitivity of the reconstruction error to different framings and parameters of the model. What we are more interested in is checking if the key messages are sound.

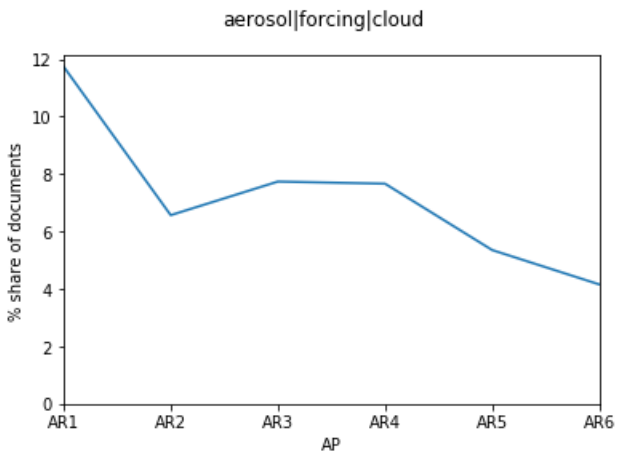
Robustness

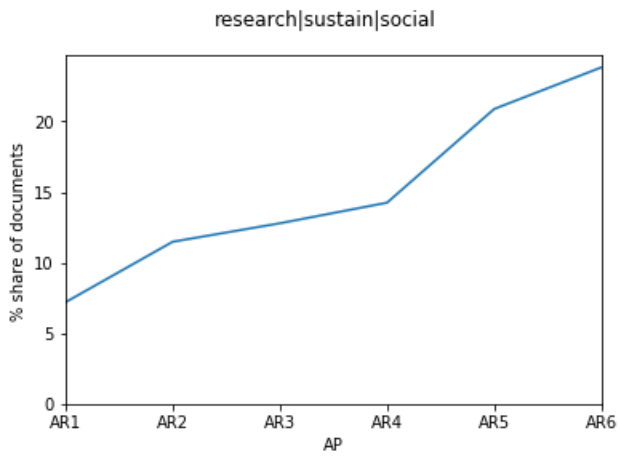
Most work focuses on the measuring the sensitivity of the reconstruction error to different framings and parameters of the model. What we are more interested in is checking if the key messages are sound.



Robustness

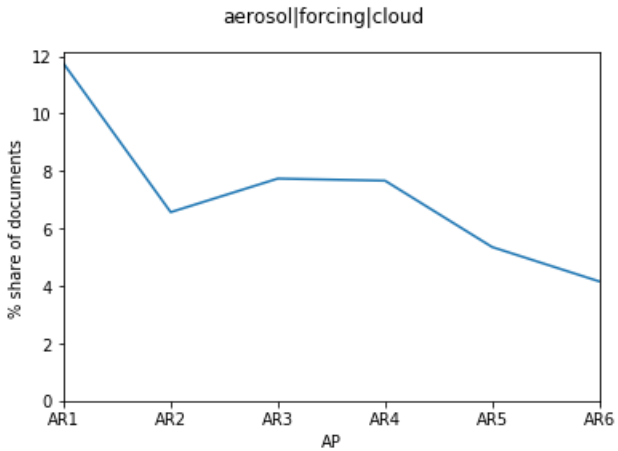
Most work focuses on the measuring the sensitivity of the reconstruction error to different framings and parameters of the model. What we are more interested in is checking if the key messages are sound.

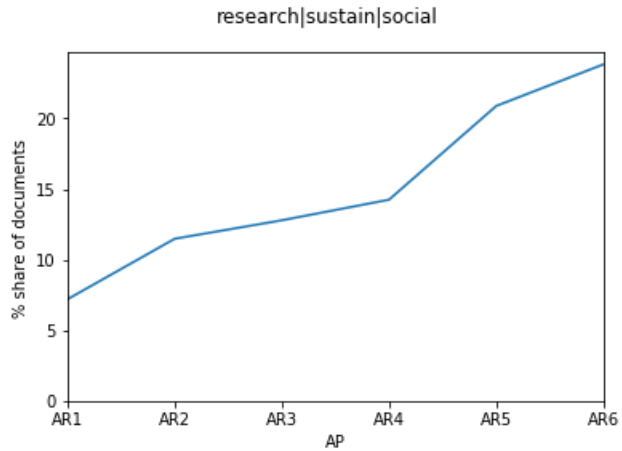


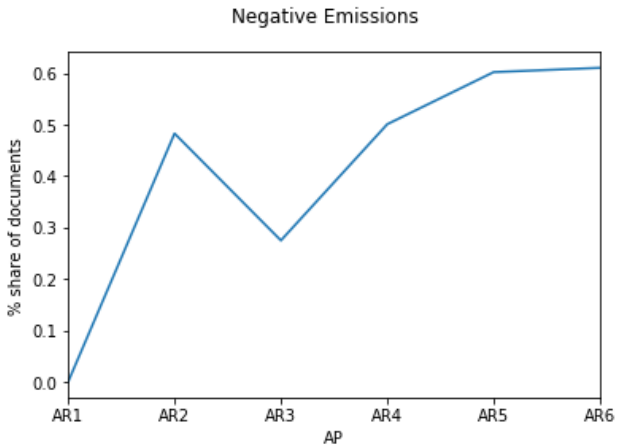


Robustness

Most work focuses on the measuring the sensitivity of the reconstruction error to different framings and parameters of the model. What we are more interested in is checking if the key messages are sound.

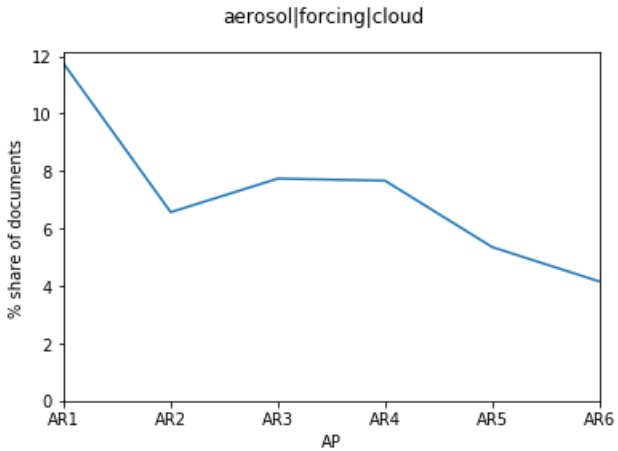


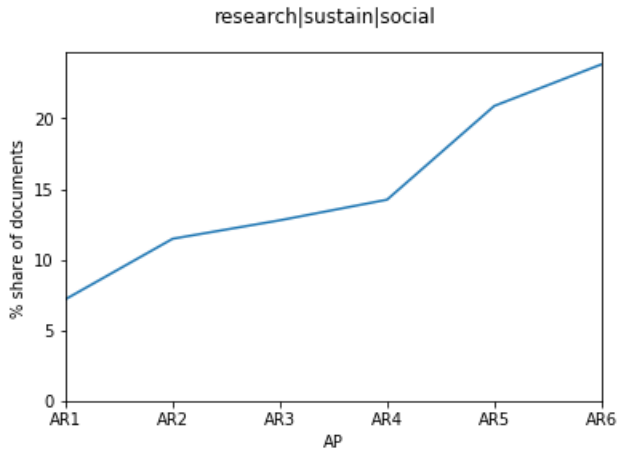


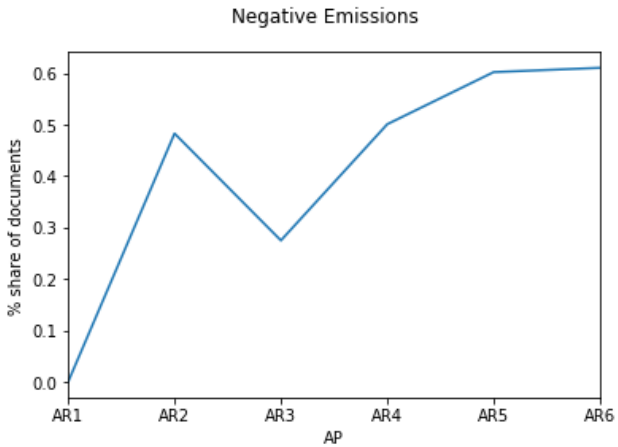


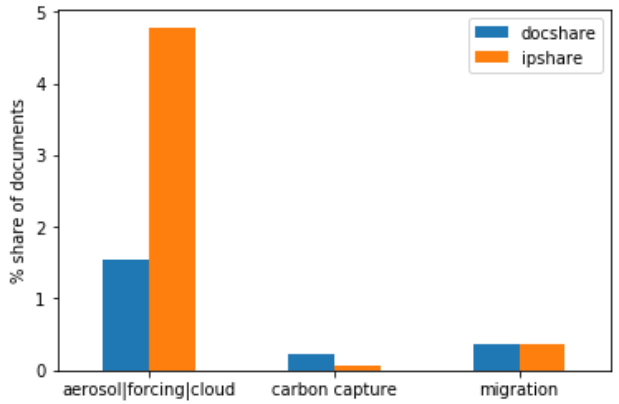
Robustness

Most work focuses on the measuring the sensitivity of the reconstruction error to different framings and parameters of the model. What we are more interested in is checking if the key messages are sound.









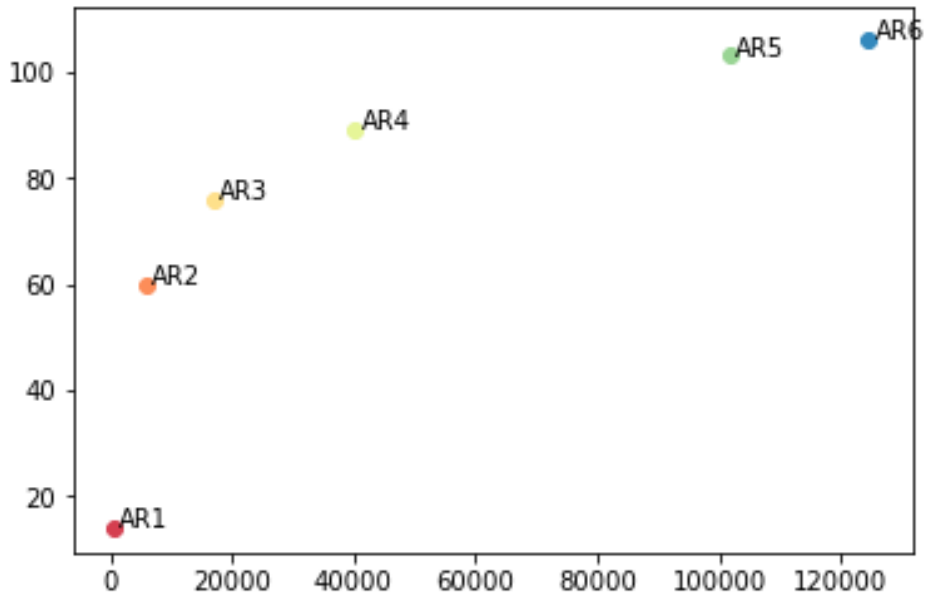
Dynamic NMF - application to climate change

- 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