| | AR1 | AR2 | AR3 | AR4 | AR5 | AR6 |
|---|---|---|---|---|---|---|
| **Years** | 1986-1989 | 1990-1994 | 1995-2000 | 2001-2006 | 2007-2013 | 2014- |
| **Documents** | 1,167 | 8,539 | 21,716 | 38,750 | 134,413 | 201,606 |
| **Words** | 2000 | 12480 | 23346 | 34637 | 71867 | 94746 |
| **New words** | change (560) | oil (287) | downscaling (217) | sres (234) | biochar (1791) | mmms (313) |
| | climate (428) | deltac (283) | degreesc (187) | petm (95) | redd (1113) | cop21 (234) |
| | co2 (318) | whole (256) | ncep (130) | amf (88) | cmip5 (679) | c3n4 (214) |
| | climatic (289) | tax (254) | fco (107) | sf5cf3 (86) | cmip3 (587) | sdg (187) |
| | model (288) | landscape (249) | pfc (98) | clc (81) | mofs (299) | zika (182) |
| | atmospheric (281) | alternative (243) | otcs (98) | embankment (81) | sdm (297) | ndcs (168) |
| | effect (280) | availability (242) | dtr (95) | cwd (79) | mof (275) | indc (164) |
| | global (224) | life (239) | nee (89) | etm (75) | biochars (252) | indcs (134) |

Table 1: Growth of Literature on Climate Change. A glossary of acronyms is provided in SI

# A Topography of Climate Change Research

Max Callaghan[1,2]

[1]Mercator Research Institute on Global Commons and Climate Change, Torgauer Straße, 10829 Berlin, Germany

[2]School of Earth and Environment, University of Leeds, Leeds LS2 9JT, United Kingdom

Draft current May 23, 2019

**The massive expansion of scientific literature on climate change challenges the Intergovenmental Panel on Climate Change (IPCC)'s ability to assess the science according to its objectives. Moreover, the number and variety of papers hinders researchers of the science-policy interface from making objective judgements about those IPCC assessments. In this paper, we present a novel application of a machine-reading approach to model the topical content of papers on climate change. This dynamic topic model provides the basis for a *topography* of climate change literature. The thematic development of the field is outlined and used to inform an analysis of the topics which are better and less well covered by IPCC reports.**

We know that the scientific literature on climate change is growing rapidly [1], and that this growth poses problems for the IPCC. Their task of providing a comprehensive and transparent summary of the literature on climate change is severely challenged by the growth of the literature [2]. Despite this, we know little about thematic trends within the literature, or how the IPCC performs in reflecting this growing body of knowledge on climate change. Understanding these trends is a crucial task if we are to assess and improve comprehensiveness in global environmental assessments.

Table 1 depicts the scale of the challenge to the IPCC. In the years since the publication of AR5, as much literature has been published as in the 30 years previously. Moreover, not only are more articles being published, but the vocabulary of climate knowledge has expanded. While the 8,539 documents published in AR2 contained 12,480 unique words, the 201,606 documents published in AR6 contain a vocabulary of 94,746 unique words.

The zika virus, the sustainable development goals, intended nationally determined contributions and mixed matrix membranes are all significant parts of the literature since 2014 which were simply not discussed in the context of climate change before the last IPCC report. In the context of this expanding vocabulary, this study employs topic modelling to draw out patterns in the content of scientific literature. Topic modelling is an unsupervised machine-learning technique, where patterns of word co-occurrences in documents are used to learn a set of topics which can be used to describe the corpus [3], and it is applied here for the first time to the whole scientific literature on climate change.

## A Topography of Climate Change

Topics, from the greek word "topos" (meaning place), refer here to concepts or themes within the literature. In this sense, the topographic map shown in figure 1 *situates* the 400,000 documents about climate change in a topical landscape derived from the 140 topics discovered through topic modelling. Using t-distributed stochastic network embedding (t-SNE)[4] as a dimensionality reduction technique, the 140 dimensional topic space is *projected* onto two dimensions, such that documents with similar topical content are placed next to each other.

## Social Sciences in IPCC citations

It was argued after the fifth assessment report that the IPCC needs to do more to incorporate knowledge from the social sciences [5], and a scientometric study of the the third assessment report claimed the IPCC gave a greater *emphasis* to natural sciences and, within the social sciences, to economics. The study, however, operationalised disciplinary emphasis as simply the number of citations from each field. Here we look at *representativeness*, that is, the share of IPCC citations in each field divided by the share of all climate related documents in that field.

Looked at this way, we see that the social sciences were indeed under-represented in earlier assessment reports, but by the fifth assessment report were over-represented. Although economics was previously over-represented among social sciences, while others were under-represented, more subdisciplines of the social sciences are now over-represented.

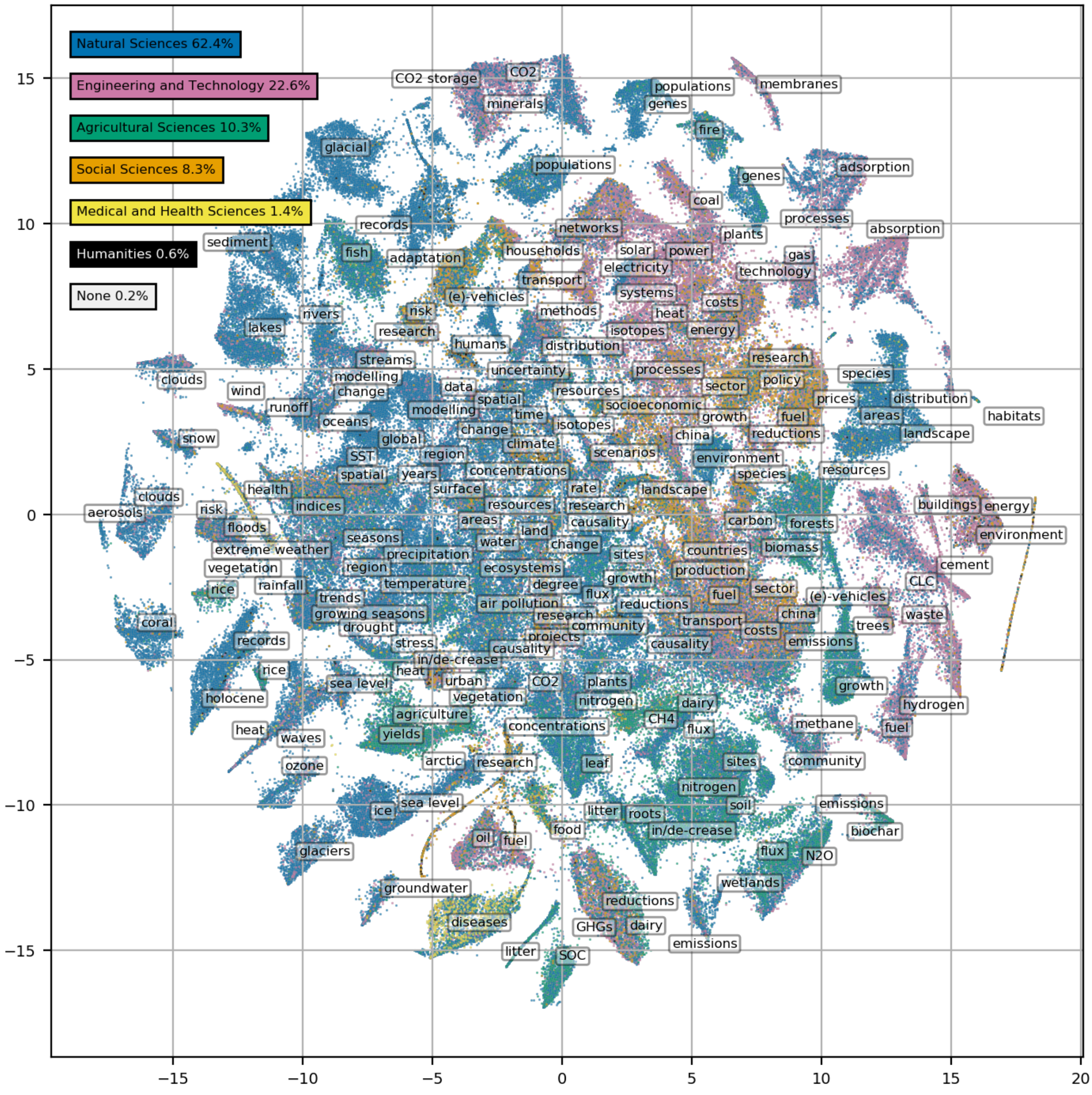


Figure 1: A map of the literature on climate change. Document positions are obtained by reducing the topic scores to two dimensions via t-SNE Documents are coloured by web of science discipline category. Topic labels are placed in the center of each of the large clusters of documents associated with each topic.

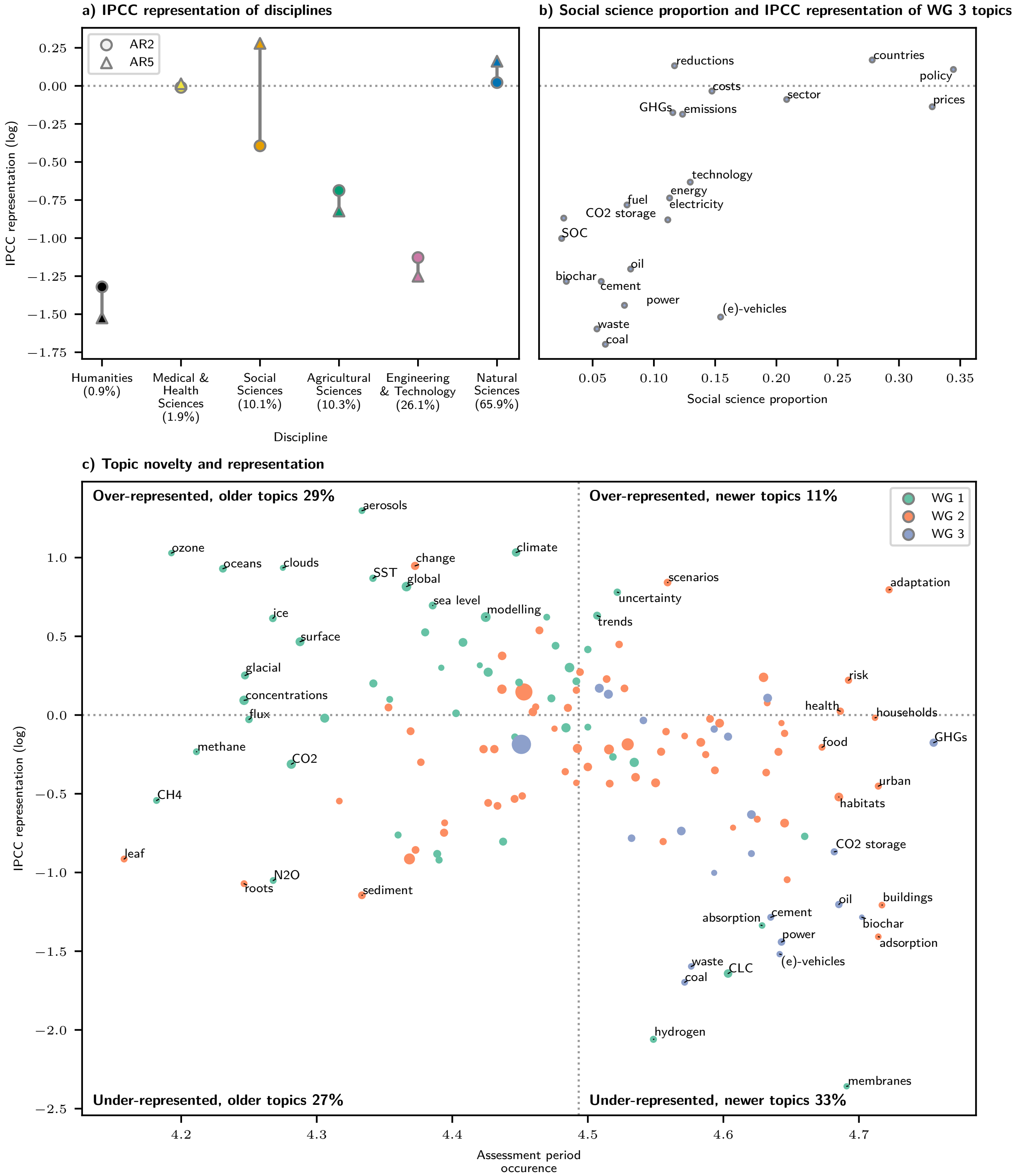


Figure 2: **a)** the representation of each discipline in IPCC reports, **b)** the representation and social science proportion of WG 3 topics, **c)** The representation and novelty of all topics

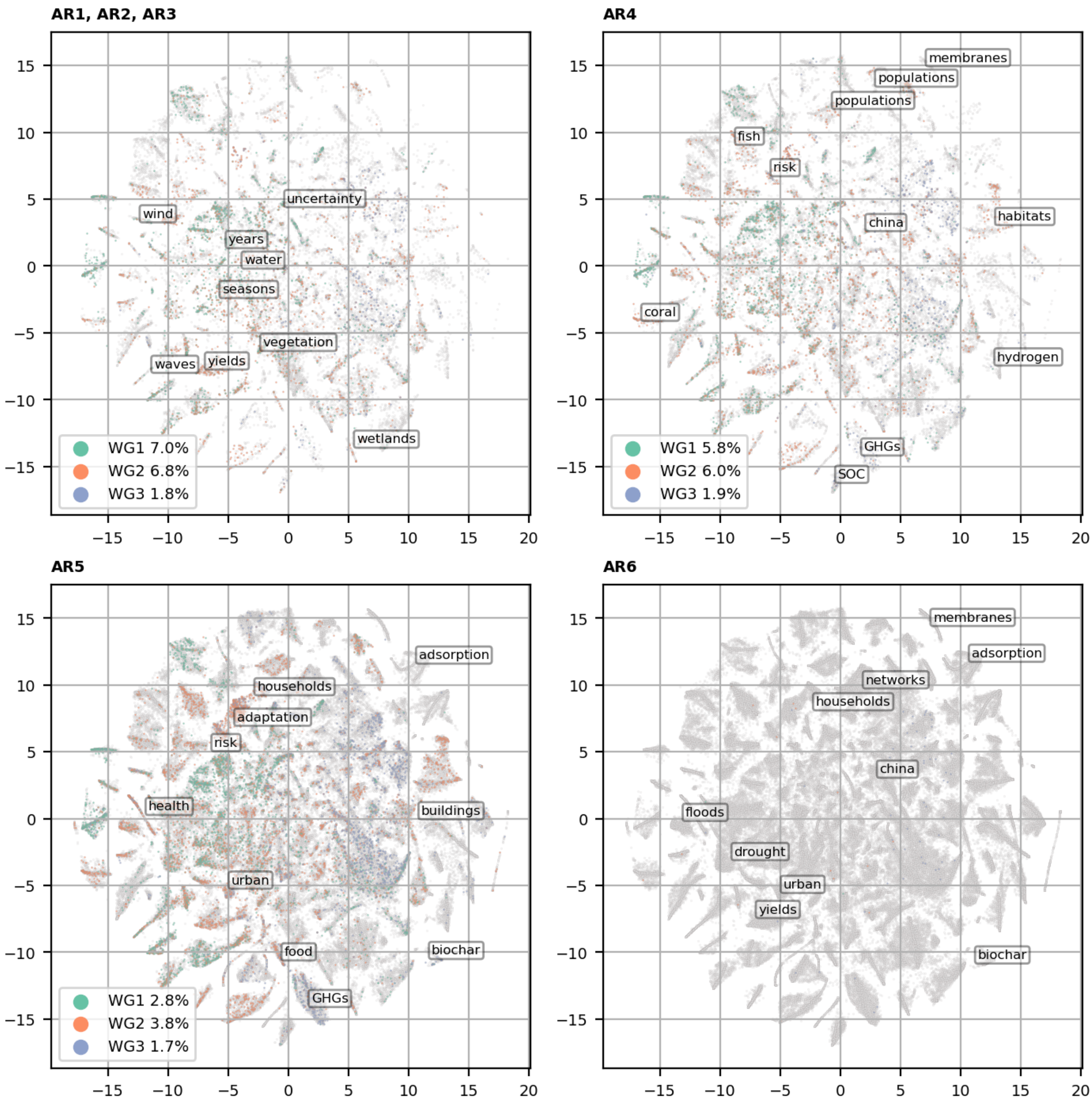


Figure 3: Evolution of the landscape of climate change literature

## References

[1] Michael Grieneisen and Minghua Zhang. The Current Status of Climate Change Research. *Nature Climate Change*, 1:72–73, 2011.

[2] Jan C. Minx, Max Callaghan, William F. Lamb, Jennifer Garard, and Ottmar Edenhofer. Learning about climate change solutions in the IPCC and beyond. *Environmental Science & Policy*, 2017.

[3] David Blei, Lawrence Carin, and David Dunson. Probabilistic topic models. *Communications of the ACM*, 55(4):77–84, 2012.

[4] Laurens van der Maaten and Geoffrey Hinton. Visualizing Data using t-SNE. *Journal of Machine Learning Research*, 9:2579–2605, 2008.

[5] David G. Victor. Embed the social sciences in climate policy - David Victor. *Nature*, 520:7–9, 2015.