Supporting Information

"Fast growing research on negative emissions"

Jan C. Minx[1,2,*], William F. Lamb[1], Max W. Callaghan[1], Lutz Bornmann[3] and Sabine Fuss[1]

[1] Mercator Research Institute on Global Commons and Climate Change, Torgauer Straße 12-15, 10829 Berlin, Germany
[2] Hertie School of Governance, Friedrichstrasse 180, 10117 Berlin, Germany
[3] Division for Science and Innovation Studies, Administrative Headquarters of the Max Planck Society, Hofgartenstr. 8, 80539 Munich, Germany
[*] corresponding author: minx@mcc-berlin.net

Scientometrics in climate change research

Scientometric methods (Leydesdorff and Milojević 2015) have been increasingly applied to analyze the evolution and contributions of climate change research. Studies focusing on the entire spectrum of the field show an exponential growth in its literature (Stanhill 2001; Grieneisen and Zhang 2011; Li, Wang et al. 2011; Haunschild, Bornmann et al. 2016; Minx, Callaghan et al. 2016). This growth is considerably faster than the average growth across all fields in the Web of Science (WoS) (Haunschild, Bornmann et al. 2016) – a subscription-based scientific citation indexing service that provides a comprehensive citation search. Currently the number of climate change publications doubles every 5-6 years in WoS, and surpassed 220,000 studies by the end of 2015 (Haunschild, Bornmann et al. 2016; Minx, Callaghan et al. 2016). Other scientometric studies have focused on specific problem areas, such as carbon taxation (Zhang, Wang et al. 2016), vulnerability (Wang, Pan et al. 2014) or authorship and disciplinary patterns in the IPCC reports (Bjurström and Polk 2011; Vasileiadou, Heimeriks et al. 2011; Corbera, Calvet-Mir et al. 2015).

Two further studies have applied scientometric methods to the field of NETs (Belter and Seidel 2013; Oldham, Szerszynski et al. 2014), although these are written in the wider context of climate engineering and therefore also address a set of solar radiation management (SRM) technologies. Both find an exponentially growing literature at low levels: by 2012 a total of about 800 publications deal with the various climate engineering technologies, predominantly ocean fertilization, SRM and land-based methods. Annual publications track at about 150 in 2012. Interestingly, with growing numbers of publications on the topic of geoengineering, Oldham et al. (2014) identify a parallel growth in patent applications related to climate engineering technologies.

We limit our analysis to the literature on negative emissions technologies, because of the distinct and important role they have in staying within the 1.5°C and (potentially) 2°C carbon budgets in the long-run and their complementarity with other mitigation options. More generally, meshing discussions on carbon dioxide removal with those on solar radiation management is often not instructive due to the very different roles played by these technology clusters in climate change mitigation (see, Clarke, Jiang et al. 2014; Fuss, Canadell et al. 2014). We also further extend the analysis to 2016, allowing us to capture a large body of new literature in a field that is exponentially growing.

Literature Search

The data for this analysis is derived from a WoS literature query up to 2016. Our search is a combination of eight strings, each comprising one of the NET technologies under study, as well as a group of generic keywords for NET research. The search string was built up iteratively to include the relevant synonyms for each technology and to exclude keywords that confounded our results. For

instance, in the case of ocean fertilization and enhanced weathering, we found that these broad terms by themselves would sample papers on ancient climatic change not directly relevant to our study, which deals only with anthropogenic climate change. Consequently, keywords such as 'ice' (from 'ice-core') and 'paleo' were found to adequately remove these without a loss of relevant studies. Note that abbreviations such as BECCS are risky search terms and often deliver unexpected results. In the case of afforestation, biochar and soil carbon management – large subject areas that do not exclusively deal with climate change mitigation – our approach was to identify only papers that directly referenced carbon dioxide sequestration. Throughout this procedure, random samples of the dataset were taken to ensure at least 90% of the queried papers met our standard for inclusion. Objectivity and reproducibility of this assessment was secured by ensuring consistency of judgements across independent reviewers. All document types were included in the search. Below we provide the full search query (which can be pasted into the "advanced search" function in Web of Science), as well as a disaggregated query for each technology.

1. Full search query

(TS = (biochar* AND ((carbon OR CO2) NEAR/3 (sequest* OR storage OR stock OR accumulat* OR capture))) OR TS = (ocean NEAR/5 iron NEAR/5 (fertili*ation OR enrichment) NOT natural NOT ice* NOT glaci*) OR TS = ((soil NEAR/3 (carbon OR CO2) NEAR/3 (sequest* OR storage)) AND ("climate change" OR "global warm*") AND (manag* OR practice* OR restoration OR land-use)) OR TS = ((afforestation OR reforestation) AND ((carbon OR CO2) NEAR/3 (sequest* OR storage))) OR (TS = (("ocean liming") AND (removal OR storage) AND (CO2 OR carbon*)) OR TS = ((geoengineer*) AND (silicate OR olivine OR albite OR CACO3)) OR TS = ((silicate OR olivine OR albite OR CACO) AND (mitigat* NEAR/3 ("climate change" OR "global warming"))) OR TS = (("ocean alkalini*") AND (remov* OR storage OR mitigat* OR sequest*) AND (CO2 OR carbon*)) OR TS = (((enhance* OR artificial*) NEAR/2 weathering) AND ((carbon OR CO2 OR "climate change" OR "global warming") NEAR/3 (remov* OR sequest* OR storage OR sink OR mitigat* OR reduc*)))) NOT TS = (glaci* OR ice* OR ordovic* OR Aptian OR Cenozo* OR Paleo* OR Mezoso*) OR (TS = (((capture OR extraction OR absorbtion) NEAR/3 (air OR atmosph*)) AND (ambient OR "atmosph* pressure*") AND (CO2 OR carbon)) OR TS = (((captur* OR extract) NEAR/3 (direct* OR "carbon dioxide") NEAR/3 (air OR atmosph*)) AND (CO2 OR carbon)) OR TS = ((*sorbent OR amine) AND capture AND (carbon OR CO2) AND ("ambient air")) OR TS = ((captur* NEAR/3 CO2 NEAR/3 (air OR atmosph*)) AND solar)) NOT TS = (phenolic OR PCB* OR particulate OR NOx OR isotope OR "heat pump" OR polycyclic OR *bacteria* OR lignin OR sink OR pollution OR photosynth* OR biofuel* OR sugar) OR TS = (BECCS OR ((biomass OR bioenerg*) AND ("CCS" OR "Carbon capture and Storage" OR "Carbon dioxide capture and Storage" OR "CO2 capture and storage")) NOT "co-fir*" NOT "co-generat*" NOT cogeneration NOT coal) OR TS = ((seagrass OR mangrove* OR macroalgae OR "blue carbon") AND ((carbon OR CO2) NEAR/3 (sequest* OR accumulat* OR storage OR capture)) AND (deforest* OR afforest* OR conserv* OR restor* OR manag*)) OR (TS = ((CDR AND (CO2 OR carbon*)) OR "negative carbon dioxide emission*" OR "negative CO2 emission*" OR "negative GHG emission*" OR "negative greenhouse gas emission*" OR "carbon-negative emission*" OR ("negative emission*" NEAR/10 carbon) OR ("negative emission*" NEAR/10 CO2)) OR TS = (geoengineering AND ((carbon OR CO2) NEAR/3 (sequest* OR accumulat* OR storage OR capture))) OR TS = (("geoengineering" OR "climate engineering") AND CDR)) NOT TS = (N2O OR nitrogen OR NOX)) NOT TS = ("bioactive equivalent combinatorial components" OR "bandwidth-efficient-channel-coding-scheme" OR "bronchial epithelial cell cultures" OR "california current system" OR comet OR mars OR exoplanet* OR "competition chambers" OR gastric OR (mercury NEAR/3 capture) OR (image NEAR/3 capture) OR "canary current system" OR "heavy metal" OR eicosanoid OR "companion cells" OR "calcium carbonate sand" OR "copper chaperone" OR "commercial cane sugar" OR "Cindoxin reductase" OR

"coupled dissolution reprecipitation" OR "carbon dioxide reforming" OR rats OR "complementarity determining regions" OR deoxycytidine) NOT PY = 2017

1.1 Soil carbon sequestration

TS = ((soil NEAR/3 (carbon OR CO2) NEAR/3 (sequest* OR storage)) AND ("climate change" OR "global warm*") AND (manag* OR practice* OR restoration OR land-use))

Soil carbon sequestration is a large subject area with many potentially relevant studies for negative emissions - a basic query for soil carbon sequestration yields ~3000 studies. We refine this search to address two issues: (1) many studies do not deal with climate change mitigation, but are primarily focused on the management of soil carbon to enhance fertility, biodiversity, and so forth; (2) many studies are focused on natural developments and phenomena related to soil carbon sequestration, rather than the purposeful management of soil carbon to mitigate carbon emissions. Arguably, the above studies may be regarded as relevant for a systematic review of soil carbon sequestration, but in our scientometric study they are excluded where possible using synonyms for "climate change" and "management" or "practices", such that the focus is primarily on the mitigation aspects of this field.

1.2 Afforestation/reforestation

TS = ((afforestation OR reforestation) AND ((carbon OR CO2) NEAR/3 (sequest* OR storage)))

Afforestation/reforestation is a very large subject area – a basic query for these two words returns ~7,600 papers. Again our strategy is to focus only on the carbon sequestration aspects of afforestation and reforestation, however since the intentional management of forests for climate change mitigation is implicit in these keywords, it was not necessary to further refine the search as in soil carbon sequestration.

1.3 Biochar

TS = (biochar* AND ((carbon OR CO2) NEAR/3 (sequest* OR storage OR stock OR accumulat* OR capture)))

The biochar query is relatively straightforward, since it refers to a distinct technology and process that is primarily oriented towards the purpose of climate change mitigation. Nonetheless, as with other technologies, the search is refined to focus on carbon sequestration and its synonyms.

1.4 Ocean fertilisation

TS = (ocean NEAR/5 iron NEAR/5 (fertili*ation OR enrichment) NOT natural NOT ice* NOT glaci*)

The literature on ocean fertilisation is relatively small, but it shares similar concepts and keywords with other fields - in particular, earth-system modelling of historical climates. Therefore we set exclusions to filter these studies out where possible.

1.5 Enhanced weathering

(TS = (("ocean liming") AND (removal OR storage) AND (CO2 OR carbon*)) OR TS = ((geoengineer*) AND (silicate OR olivine OR albite OR CACO3)) OR TS = ((silicate OR olivine OR albite OR CACO) AND (mitigat* NEAR/3 ("climate change" OR "global warming"))) OR TS = (("ocean alkalini*") AND (remov* OR storage OR mitigat* OR sequest*) AND (CO2 OR carbon*)) OR TS = (((enhance* OR artificial*) NEAR/2 weathering) AND ((carbon OR CO2 OR "climate change" OR "global warming") NEAR/3 (remov* OR sequest* OR storage OR sink OR mitigat* OR reduc*)))) NOT TS = (glaci* OR ice* OR ordovic* OR Aptian OR Cenozo* OR Paleo* OR Mezoso*)

As with ocean fertilisation, enhanced weathering involves the anthropogenic enhancement of a natural earth-system process. While palaeo-climate studies may indeed be important for a full picture of this technology, our focus remains on the literature dealing with contemporary applications of enhanced weathering for carbon sequestration; this informs our relatively narrow search criteria and extensive exclusion list.

1.6 Bioenergy carbon capture and storage

TS = (BECCS OR ((biomass OR bioenerg*) AND ("CCS" OR "Carbon capture and Storage" OR "Carbon dioxide capture and Storage" OR "CO2 capture and storage")) NOT "co-fir*" NOT "co-generat*" NOT cogeneration NOT coal)

There are two main challenges in searching for the literature on BECCS: (1) much of the literature is not strictly carbon negative, i.e. it focuses on fossil-fuel based CCS, or co-generation with biomass; (2) the acronym “CCS” is widely used in other fields, particularly the health sciences. While the latter issue can be addressed with a list of exclusions (see section 1.9), delineating between biomass and non-biomass based CCS is not trivial. Our procedure is to set exclusions for co-firing, co-generation and coal. We recognise that these technologies are similar and bear relevance for one another, but as with other parts of the search query we tend towards exclusion rather than inclusion in defining the final literature corpus.

1.7 Direct air capture

(TS = (((capture OR extraction OR absorbtion) NEAR/3 (air OR atmosph*)) AND (ambient OR "atmosph* pressure*") AND (CO2 OR carbon)) OR TS = (((captur* OR extract) NEAR/3 (direct* OR "carbon dioxide") NEAR/3 (air OR atmosph*)) AND (CO2 OR carbon)) OR TS = ((*sorbent OR amine) AND capture AND (carbon OR CO2) AND ("ambient air")) OR TS = ((captur* NEAR/3 CO2 NEAR/3 (air OR atmosph*)) AND solar)) NOT TS = (phenolic OR PCB* OR particulate OR NOx OR isotope OR "heat pump" OR polycyclic OR *bacteria* OR lignin OR sink OR pollution OR photosynth* OR biofuel* OR sugar)

Identifying direct air capture studies is non-trivial. There are many related processes and synonyms, such as the direct air capture of non-CO2 pollutants, or the natural capture of CO2 from air via photosynthesis. We have therefore constructed a very detailed query with an extensive exclusion list, relying on existing literature reviews to narrow down our search (Sanz-Perez et al., 2016). This includes studies that examine post-combustion CO2 capture, as these are often discussed in parallel, or provide benchmarking and comparison to direct air capture.

1.8 Blue carbon

TS = ((seagrass OR mangrove* OR macroalgae OR "blue carbon") AND ((carbon OR CO2) NEAR/3 (sequest* OR accumulat* OR storage OR capture)) AND (deforest* OR afforest* OR conserv* OR restor* OR manag*))

The blue carbon query is similar to that of afforestation/reforestation, but is more specific to coastal mangrove and seagrass ecosystems. As with the other land-based NETs, it is restricted to: (1) the sequestration of carbon; and (2) the active management, restoration or conservation of these systems.

1.9 Generic synonyms

(TS = ((CDR AND (CO2 OR carbon*)) OR "negative carbon dioxide emission*" OR "negative CO2 emission*" OR "negative GHG emission*" OR "negative greenhouse gas emission*" OR "carbon-negative emission*" OR ("negative emission*" NEAR/10 carbon) OR ("negative emission*" NEAR/10 CO2)) OR TS = (geoengineering AND ((carbon OR CO2) NEAR/3 (sequest* OR accumulat* OR storage OR capture))) OR TS = (("geoengineering" OR "climate engineering") AND CDR)) NOT TS = (N2O OR nitrogen OR NOX)

The generic synonyms are terms and phrases that tend to be used in high level discourses on negative emissions (e.g. the climate policy literature) but may not be captured by the individual technology searches. Note that we still have to set exclusions for papers on negative (non-CO2) GHGs, as these do not have a sequestration component. The inclusion of "CDR" also generates a number of unrelated studies which use the same acronym – these are dealt with in the general exclusions (section 1.9).

1.9 General exclusions

NOT TS = ("bioactive equivalent combinatorial components" OR "bandwidth-efficient-channel-coding-scheme" OR "bronchial epithelial cell cultures" OR "california current system" OR comet OR mars OR exoplanet* OR "competition chambers" OR gastric OR (mercury NEAR/3 capture) OR (image NEAR/3 capture) OR "canary current system" OR "heavy metal" OR eicosanoid OR "companion cells" OR "calcium carbonate sand" OR "copper chaperone" OR "commercial cane sugar" OR "Cindoxin reductase" OR "coupled dissolution reprecipitation" OR "carbon dioxide reforming" OR rats OR "complementarity determining regions" OR deoxycytidine)

These exclusions are primarily to remove unexpected results from the acronyms "CCS" and "CDR", as well as astronomical studies on extra-terrestrial carbon cycles. Note that a combination of technology-specific exclusions have been applied, as well as this general list. The general exclusions are not relevant in any context for our query (they tend to be in the health sciences), whereas technology-specific exclusions are not relevant in that particular context, but may be in others (hence photosynth* is excluded in direct air capture, but not in afforestation).

Topic modelling

As well as understanding the growth in NET publications and its distribution across different subject categories, it is interesting to identify and analyse the underlying topical themes of this body of literature. We might ask, for instance, which themes pervade the entire collection of documents, and how they relate to one another. Since it is impractical to perform this task manually in large collections of papers, we apply a probabilistic topic model called Latent Dirichlet Allocation (LDA) to discover these themes. In a recent paper, Boussalis and Coan (2016) apply this technique to identify common discourses in climate change denial.

LDA is an increasingly common tool for analysing large collections of text (Blei, Ng et al. 2003; Blei 2012). It proceeds on the assumption that each document in a collection is a combination of topics in different proportion; and that each topic contains a certain distribution of words. We use a variational expectation-maximization (VEM) algorithm to estimate this hidden topic structure based on the observed words in each document. LDA generates a list of topics with the words that constitute them at given probabilities, and labels each document with the probability that it belongs to each topic. Assuming that abstracts adequately represent the major contents of their underlying papers, we apply the R package topicmodels (Grün and Hornik 2011) to the abstracts, titles and keywords of papers only.

The number of topics in an LDA needs to be specified a priori. As in other clustering and latent variable methodologies, this is a non-trivial step in the analysis: too few topics and their individual word distributions will be insufficiently precise to classify; too many topics and the analysis will be overly complex and difficult to interpret. There is some debate around the optimal number, although it is generally assumed that quantitative methods should be complemented by human judgement and validation (Chang, Boyd-Graber et al. 2009; Wallach, Murray et al. 2009; Zhao, Chen et al. 2015; Boussalis and Coan 2016). Accordingly, our procedure was to sample different number of topics (corresponding to the approximate 'elbow' point of a log likelihood plot) and then manually analyze the resulting word distributions. The highest correlating documents for each topic provided important context for this exercise and allows us to identify consistent themes across the samples, as well as semantically meaningless topics. Overall, 19 topics proved to be a meaningful and manageable number, covering a broad spectrum of themes, while minimizing uninterpretable results.

Normalisation of citation impact

The citation data for the publications in our set are from an in-house database at the Max Planck Society which is based on WoS. For every publication, the citation window for impact measurement is from publication year up to the end of 2015. To ensure fairness of comparison, we therefore do not include papers written after 2013 in Table 2 of the manuscript. Disciplines vary in their citation culture, and older papers tend to accrue a greater number of citations (Althouse, West et al. 2009). As negative emissions research is a multi-disciplinary effort with a significant number of publications in recent years, we normalize papers by discipline and year in order to compare their relative impacts to the field. The normalization procedure follows a common method in bibliometrics (Bornmann and Marx 2015): (1) each paper is assigned to a reference set of papers from the same year and WoS subject category; (2) the number of citations in each reference set is averaged; (3) the normalized citation score is calculated as the total number of citations for a given paper divided by the reference set average. For papers published in journals which are assigned to a multi-disciplinary subject category (e.g. *Nature* and *Science*), the average citation impact of this subject category is used as reference set for the normalization.

References

Althouse, B. M., J. D. West, et al. (2009). "Differences in impact factor across fields and over time." Journal of the American Society for Information Science and Technology 60(1): 27-34.

Belter, C. W. and D. J. Seidel (2013). "A bibliometric analysis of climate engineering research." Wiley Interdisciplinary Reviews: Climate Change 4(5): 417-427.

Bjurström, A. and M. Polk (2011). "Physical and economic bias in climate change research: a scientometric study of IPCC Third Assessment Report." Climatic Change 108(1): 1-22.

Blei, D. M. (2012). "Probabilistic topic models." Commun. ACM 55(4): 77-84.

Blei, D. M., A. Y. Ng, et al. (2003). "Latent dirichlet allocation." J. Mach. Learn. Res. 3: 993-1022.

Bornmann, L. and W. Marx (2015). "Methods for the generation of normalized citation impact scores in bibliometrics: Which method best reflects the judgements of experts?" Journal of Informetrics 9(2): 408-418.

Boussalis, C. and T. G. Coan (2016). "Text-mining the signals of climate change doubt." Global Environmental Change 36: 89-100.

Chang, J., J. Boyd-Graber, et al. (2009). Reading tea leaves: how humans interpret topic models. Proceedings of the 22nd International Conference on Neural Information Processing Systems. Vancouver, British Columbia, Canada, Curran Associates Inc.: 288-296.

Clarke, L., K. Jiang, et al. (2014). Assessing Transformation Pathways. Climate Change 2014: Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Edenhofer, O., R. Pichs-Madruga, Y. Sokona, E. Farahani, S. Kadner, K. Seyboth, A. Adler, I. Baum, S. Brunner, P. Eickemeier, B. Kriemann, J. Savolainen, S. Schlömer, C. von Stechow, T. Zwickel and J.C. Minx (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Corbera, E., L. Calvet-Mir, et al. (2015). "Patterns of authorship in the IPCC Working Group III report." Nature Clim. Change advance online publication.

Fuss, S., J. G. Canadell, et al. (2014). "Betting on negative emissions." Nature Clim. Change 4(10): 850-853.

Grieneisen, M. L. and M. Zhang (2011). "The current status of climate change research." Nature Clim. Change 1(2): 72-73.

Grün, B. and K. Hornik (2011). "topicmodels: An R Package for Fitting Topic Models." 2011 40(13): 30.

Haunschild, R., L. Bornmann, et al. (2016). "Climate Change Research in View of Bibliometrics." PLoS ONE 11(7): e0160393.

Leydesdorff, L. and S. Milojević (2015). Scientometrics. International Encyclopedia of the Social & Behavioral Sciences (Second Edition). Oxford, Elsevier: 322-327.

Li, J., M.-H. Wang, et al. (2011). "Trends in research on global climate change: A Science Citation Index Expanded-based analysis." Global and Planetary Change 77(1–2): 13-20.

Minx, J. C., M. W. Callaghan, et al. (2016). "Learning about climate change solutions." Working Paper.

Oldham, P., B. Szerszynski, et al. (2014). "Mapping the landscape of climate engineering." Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 372(2031).

Stanhill, G. (2001). "The Growth of Climate Change Science: A Scientometric Study." Climatic Change 48(2): 515-524.

Vasileiadou, E., G. Heimeriks, et al. (2011). "Exploring the impact of the IPCC Assessment Reports on science." Environmental Science & Policy 14(8): 1052-1061.

Wallach, H. M., I. Murray, et al. (2009). Evaluation methods for topic models. Proceedings of the 26th Annual International Conference on Machine Learning. Montreal, Quebec, Canada, ACM: 1105-1112.

Wang, B., S.-Y. Pan, et al. (2014). "An overview of climate change vulnerability: a bibliometric analysis based on Web of Science database." Natural Hazards 74(3): 1649-1666.

Zhang, K., Q. Wang, et al. (2016). "A bibliometric analysis of research on carbon tax from 1989 to 2014." Renewable and Sustainable Energy Reviews 58: 297-310.

Zhao, W., J. J. Chen, et al. (2015). "A heuristic approach to determine an appropriate number of topics in topic modeling." BMC Bioinformatics 16(13): 1-10.