Science of cloud and climate science: An analysis of the literature over the past 50 years

Sullivan Sylvia[1] and Hoose Corinna[2]

[1]University of Arizona
[2]Karlsruhe Institute of Technology

November 16, 2022

Abstract

Clouds pose a particularly difficult challenge within Earth's climate system. They are relatively small in spatiotemporal scale but still have a strong influence on radiative fluxes, global circulation, and precipitation patterns. Increasing research attention has been devoted to them over the past 50 years, and we give a summary of the resulting body of scientific literature in this introductory chapter. Articles on clouds and climate are doubling every 8 years, a rate about twice that of scientific publications generally. This expanding number of publications correlates with more citations, but citation rates have also slowed in the most recent decade, despite a growing number of atmospheric science students. We show some basic "science of science" (SciSci) analyses of the clouds and climate literature, such as authorship networks or abstract text mining for techniques, and suggest that further SciSci analyses may help us to process the proliferation of results on clouds and climate and optimize how we do research in the crucial years ahead.

Science of cloud and climate science: An analysis of the literature over the past 50 years

S. C. Sullivan[1,*] **and C. Hoose**[1]

[1]Institute of Meteorology and Climate Research, Karlsruhe Institute of Technology, Karlsruhe, Germany
*Now at Department of Chemical and Environmental Engineering, University of Arizona, Arizona, USA

Corresponding author: Sylvia Sullivan, sylvia@arizona.edu

Summary

Clouds pose a particularly difficult challenge within Earth's climate system. They are relatively small in spatiotemporal scale but still have a strong influence on radiative fluxes, global circulation, and precipitation patterns. Increasing research attention has been devoted to them over the past 50 years, and we give a summary of the resulting body of scientific literature in this introductory chapter. Articles on clouds and climate are doubling every 8 years, a rate about twice that of scientific publications generally. This expanding number of publications correlates with more citations, but citation rates have also slowed in the most recent decade, despite a growing number of atmospheric science students. We show some basic "science of science" (SciSci) analyses of the clouds and climate literature, such as authorship networks or abstract text mining for techniques, and suggest that further SciSci analyses may help us to process the proliferation of results on clouds and climate and optimize how we do research in the crucial years ahead.

1 Research on Clouds and Climate

Clouds have a multi-faceted impact on the climate. They affect the terrestrial radiative balance by reflecting visible and ultraviolet radiation from the sun and absorbing infrared radiation from the Earth surface. They are also coupled to the large-scale circulation. Cloud formation is determined by where circulation patterns bring moisture aloft, while the reflection and absorption of radiation by clouds also feeds back on circulation. Finally, precipitation is generated from clouds. The intensity, frequency, and duration of precipitation are determined by cloud structure and dynamics.

Climatic impacts of clouds are also multi-scale, involving a huge range of processes, from new particle formation at the nanometer scale up to atmospheric wave propagation over thousands of kilometers. Simulating or measuring these 15 orders of magnitude challenges our computational and observational tools. And with phenomena as diverse as turbulence and nucleation, and forms as varied as cumulonimbus towers and stratocumulus decks, clouds challenge our ability to simplify or generalize.

This combination of impact and complexity has piqued the interest of ever more researchers and funding agencies and given rise to an imposing body of scientific literature. As in other scientific fields, this rapid growth in publication has motivated systematic review and meta-analysis, in this case through entities like the Intergovernmental Panel on Climate Change (IPCC) or the Climate Model Intercomparison Projects (CMIP). The fifth assessment report of the IPCC contained a chapter dedicated to clouds and climate (Boucher et al., 2013), and the most recent report devotes sections to cloud feedbacks and water cycle changes with warming (Forster et al., 2021; Douville et al., 2021). Within the most recent CMIP experimental design is an endorsed model intercomparison project focused exclusively on cloud feedbacks (CFMIP) (Webb et al., 2017).

This monograph represents one of such increasingly important community-wide review efforts. With primarily early-career lead authors, each chapter provides a review of the existing scientific literature and open questions in a given subfield. The volume is intended to act as a resource for graduate students, both to orient in their new subfield and gain fluency in related ones. For young scientists, it may direct their energy toward high-impact questions or help them formalize their future research program, and for more established scientists, it may generate ideas for collaboration.

Chapters have been organized around the climatic impacts of clouds on radiation, circulation, and precipitation described above (Fig. 1). Chapters 2 through 4 focus on the formation of liquid droplets and ice crystals on particulates and how this cloud formation can affect radiative fluxes, both at the surface and in the atmosphere. Chapters 5 through 9 review research on the most climatically-relevant cloud types including Arctic mixed-phase, extratropical, tropical marine low, and tropical organized deep. Cloud-

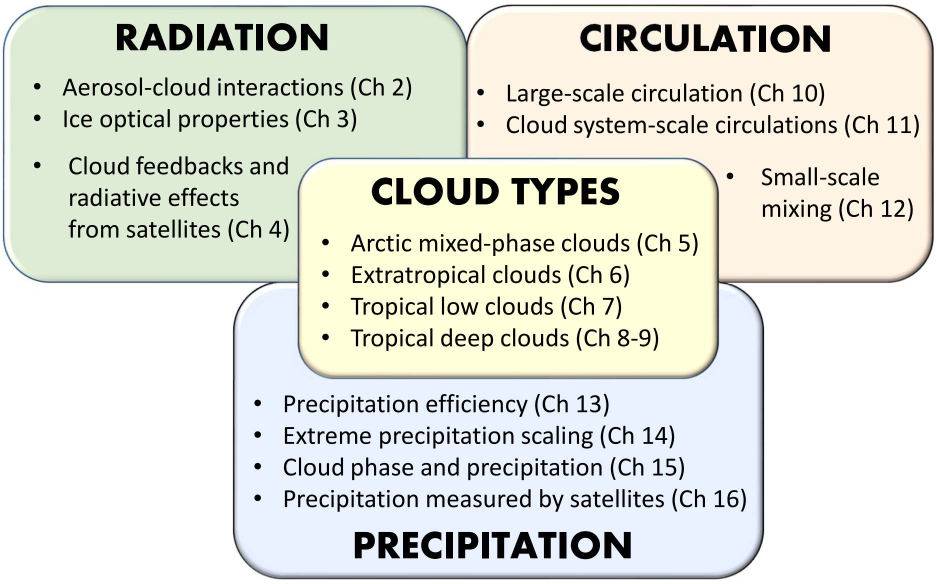

Figure 1. The monograph chapters are organized around climatic impacts of clouds on radiation, circulation, and precipitation.

circulation coupling at the global, meso, and micro scales are covered in Chapters 10 through 12. The final monograph section discusses precipitation efficiency, extremes, phase, and measurements (Chapters 13-16).

1.1 Science of Science for Clouds and Climate

As an imposing body of scientific literature might already suggest, clouds and climate research at the institute or university level has tended to answer questions by performing new simulations or analyses, rather than by synthesizing output from existing studies. There could be many possible reasons for this focus on generating new output, for example limited code and data documentation and accessibility in the past or cost and difficulty of storing and analyzing large volumes of data. In any case, systematic review at the community level is fueled by this very large number of individual studies. In this introductory chapter, we posit that analyzing how these individual studies are produced in a kind of “science of science” is another meta-exercise that would be useful for the clouds and climate community.

“Science of science” (SciSci) is an emerging area of study that combines scientometrics and the sociology of science to understand and optimize how science is done, from the emergence of new paradigms to the career path of students. As Fortunato et al. (2018) note in their review, SciSci has been driven in recent years by increasingly quantitative and accessible data on publications in databases like Scopus or Web of Science and by collaboration of natural, computational, and social scientists. Although some SciSci findings are domain- or culture-specific, a number of generalizable statements can be made. For example, networks of scientific concepts, tools, and authors tend to densify over time, indicating risk aversion and the tendency to select questions and collaborations conservatively (Fortunato et al., 2018; Foster et al., 2015). Such densification can be dangerous, as a small subset of authors cite one another and reinforce established hypotheses in an echo chamber effect.

SciSci analyses have not yet been done for clouds and climate research to our knowledge. Two decades ago, Geerts (1999) aggregated data on atmospheric science publications and found an increasing number of journals, pages per journal, and words per page in articles published between 1965 and 1995. But we have not found studies building upon this one or studies focused exclusively on climatic impacts of clouds. As a field with burgeoning student interest and the pressure to inform climate policy reliably and rapidly, clouds and climate research could benefit from such meta-study. To catalyze these efforts, we perform preliminary SciSci analyses in this introductory chapter.

1.2 Publication Data and Methods

We draw our publication data from the Scopus database of Elsevier, which contains abstracts and citations of peer-reviewed academic literature. A set of boolean keywords define the Scopus query for publications on clouds and climate, as well those filtered by theme (Tab. 1). For example for all publications in the field, we require that the title, abstract, or keywords contain both the term "clouds" and the term "climate" (TITLE-ABS-KEY("clouds" AND "climate")). We also require that the document be an English language journal article past the review stage (LIMIT TO(PUBSTAGE("final")) AND LIMIT-TO(DOCTYPE,"ar") AND LIMIT-TO(LANGUAGE,"English")) and that the journal domain be Earth or environmental sciences (LIMIT-TO(SUBJAREA("EART") OR LIMIT-TO(SUBJAREA("ENVI")). We find a non-negligible number of publications on Mars for some queries and additionally omit these (NOT TITLE-ABS-KEY("Mars")).

Scopus queries are also used to classify techniques by searching all clouds and climate abstracts for keywords. For example, we classify abstracts that contain the strings "model", "parameterization", "simulation", "GCM", "trajector", or "radiative transfer" as modelling work. Abstracts can be classified as employing multiple techniques if they satisfy multiple queries. The anonymous Author ID field of Scopus queries is used with the Python networkx package to generate authorship networks. Finally, the titles associated with some queries are used to generated word clouds, using the Python WordCloud package. WordCloud takes the title text and fills space with individual words, sized by their frequency. A set of default "stop words" is omitted (listed in `https://github.com/amueller/word_cloud/blob/master/wordcloud/stopwords`) to which we add the following words: *cloud, climate, using, change, comparison, evaluation, effect, global, study, effects, atmospheric, changes, response, based, part, characteristic,* and *influence.* By eliminating these words, concrete article themes and techniques emerge more clearly from the title text.

The queries in Tab. 1 are denoted **Clouds and Climate** for all publications on clouds and climate; **Impacts** for publications decomposed into cloud impacts on radiation, circulation, and precipitation corresponding to the monograph sections; **Chapters** for publications decomposed according to the monograph chapters; and **Techniques** for identification of the techniques used in the abstracts of the clouds and climate publications. With the queries in Tab. 1, advanced searches can be reproduced at `www.scopus.com/search/form.uri?display=advanced`. Publication data for the figures has been downloaded on 3 September 2020 and is shown only through the end of 2019. The only non-Scopus data used concerns doctorates granted and funding awarded for the United States, which we take from the National Center for Science and Engineering Statistics (NCSES) Survey Data at `https://ncsesdata.nsf.gov/home`. For US doctorate degrees granted in geosciences, we look at the Survey of Earned Doctorates at `https://ncsesdata.nsf.gov/builder/sed` with Doctorate Recipients as our measure and Academic Discipline: Geosciences, atmospheric sciences, and ocean sciences as our dimension. For US annual funding, we look at the Survey of Federal Funds for Research and Development at `https://ncsesdata.nsf.gov/builder/ffs` with Research Obligations as our measure and Fields of Study: Atmospheric Science as our dimension.

Table 1. Boolean syntax for Scopus database queries. We filter for English-language journal articles ("ar") in the finalized publication stage ("final"), classified as Earth and Planetary Sciences ("EART") or Environmental Sciences ("ENVI"). Filters other than TITLE-ABS-KEY are held constant for all searches and denoted *other filters*.

	Queries
Clouds and Climate	TITLE-ABS-KEY("clouds" AND "climate") AND NOT TITLE-ABS-KEY("Mars") AND LIMIT-TO(PUBSTAGE("final")) AND LIMIT-TO(DOCTYPE,"ar") AND LIMIT-TO(SUBJAREA,"EART") OR LIMIT-TO(SUBJAREA,"ENVI") AND LIMIT-TO(LANGUAGE,"English")
Impacts + *other filters*	TITLE-ABS-KEY("clouds" AND "climate" AND "circulation") TITLE-ABS-KEY("clouds" AND "climate" AND "radiati*") TITLE-ABS-KEY("clouds" AND "climate" AND "precipitation")
Chapters + *other filters*	TITLE-ABS-KEY("equilibrium climate sensitivity" OR "cloud feedback*") TITLE-ABS-KEY(("radiative transfer" AND "climate" AND "cloud") OR "cloud radiative effect*") TITLE-ABS-KEY(("cloud classification") TITLE-ABS-KEY(("cloud microphysics" OR ("cloud" AND "microphysics") OR "microphysics parameterization") TITLE-ABS-KEY(("aerosol indirect effect" OR "aerosol-cloud interaction") TITLE-ABS-KEY(("atmospher*" AND ("dynamical core" OR "primitive equations")) TITLE-ABS-KEY("radiative-convective equilibrium" OR "convective organization" OR "convective aggregation" OR "organized convection") TITLE-ABS-KEY("cloud-circulation coupling" OR ("cloud" AND "large-scale circulation")) TITLE-ABS-KEY("cloud" AND "field campaign") TITLE-ABS-KEY("cloud" AND "ground-based measurement") TITLE-ABS-KEY(("cloud" AND "machine learning") OR ("cloud" AND "causal inference"))
Techniques	ABS("in-situ" OR "flight" OR "campaign" OR "aircraft" OR "rocket" OR "drone") ABS("model" OR "parameterization" OR "simulation" OR "GCM" OR "trajector" OR "radiative transfer") ABS("reanalys" OR "emission") ABS("satellite" OR "CERES" OR "TRMM" OR "ISCCP" OR "remote sens" OR "retrieval" OR "imager" OR "CALIPSO" OR "CloudSat" OR "MODIS" OR "mission") ABS("ground-based" OR "station" OR "meteorological observator" OR "rain gauge" OR "site" OR "SHEBA" OR "flux tower") ABS("laboratory" OR "chamber" OR "chemical characteriz") ABS("observation")

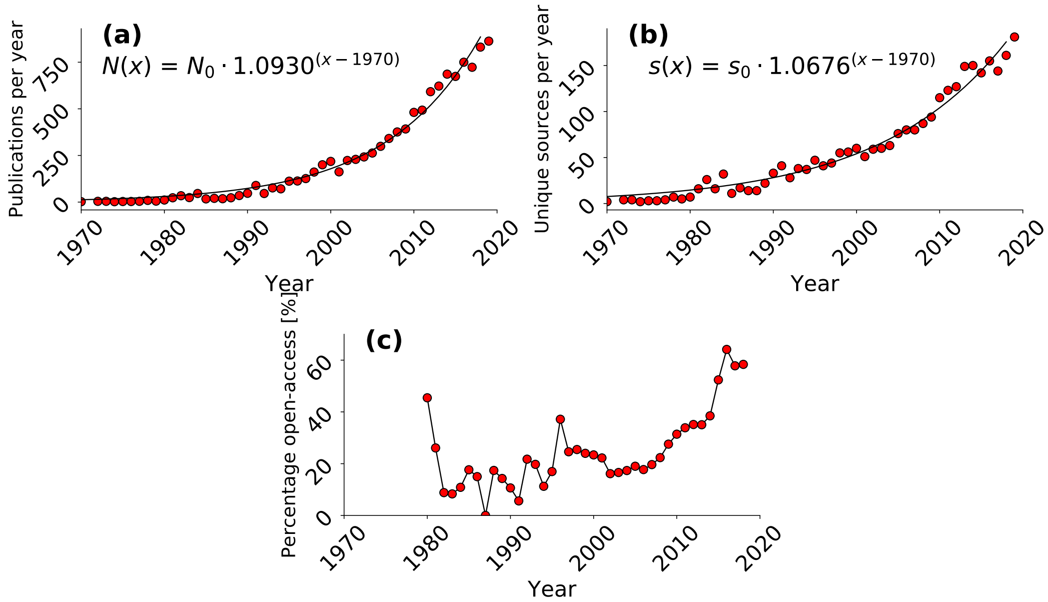

Figure 2. Atmospheric science publications, particularly open-access ones, and their unique sources have all increased dramatically over the past 50 years. Total number of publications per year from 1970 through 2019 with the **Clouds and Climate** query in Tab. 1 and their exponential fit (panel a). Number of unique journals for these publications per year from 1970 through 2019 and their exponential fit (panel b). Percentage of publications that have been open-access over time (panel c).

2 Publications on Clouds and Climate

We quantify the rapid growth of interest in clouds and climate with the output of the Scopus **Clouds and Climate** query. We find that the number of articles published on clouds and climate has been doubling every 8 years since 1970 (Fig. 2a). While an average of 27 articles per year were published in the 1980s, this rate had increased more than ten-fold by the 2000s. In the 2010s, an average of 660 articles per year were published. By comparison, Price (1963) and Fortunato et al. (2018) find a 15-year doubling period for scientific articles more generally, while Milojević (2015) cites 16- and 19-year doubling periods for physics and biomedical articles respectively.

We could also consider how this growth compares to that of other subfields in climate science. Searching instead for publications on *ocean and climate* or *air pollution and climate*, these have doubling times of 7.9 and 8.5 years respectively, comparable to that of clouds and climate (not shown). Publications on the *biosphere and climate* are growing somewhat more gradually with a doubling time of 8.9 years, while publications on the *land surface*, the *cryosphere*, or the *carbon cycle* have stronger recent growth relative to cloud research, doubling in only 6.6 years, 5.1 years, and 6.0 years respectively. These doubling rates show that publication growth for climate research is about twice as fast as that for general scientific research.

Not only is publication number on clouds and climate rapidly rising, the number of unique journals has also increased over time, doubling roughly every 11 years (Fig. 2b). In the 1980s, with the focus more often on meteorology, the *Journal of Atmospheric Sciences* published the most articles. By the 1990s, spurred by the release of the Charney Report and the launch of early geostationary satellites, interest broadened to the

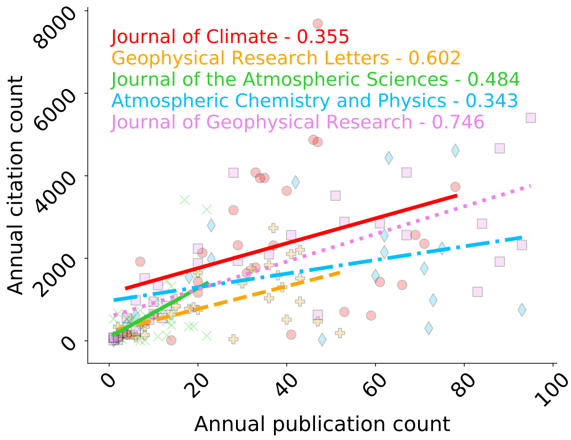

Figure 3. Publication and citation number are positively correlated, but not always strongly. Annual publication counts from 1970 through 2019 scattered against annual citation counts over the same period for the five most prolific atmospheric science publications; *Journal of Climate* and *Journal of the Atmospheric Sciences* are shown as solid traces, *Geophysical Research Letters* as a dashed trace, *Atmospheric Chemistry and Physics* as a dotted-dashed trace, and *Journal of Geophysical Research* as a dotted trace
. Correlation coefficients are given next to the publication name in the legend.

role of clouds in climate with more publications now in the *Journal of Climate* and *Journal of Geophysical Research*. Although only initiated in 2001, *Atmospheric Chemistry and Physics* (*ACP*) already published the most articles by the 2010s. As an open-access publisher, *ACP* output has also driven the large increase in open-access publication percentage over the past 4 decades from less than 10% in the 1980s to more than 50% in the late 2010s (Fig. 2c).

2.1 Citation and Readability

Does this expanding body of literature correspond to higher readership and, hence, citation? For the five most prolific journals publishing on clouds and climate, higher annual publication counts do correlate with higher annual citation counts from 1970 through 2019, but not always very strongly (Fig. 3). In spite of – or perhaps because of – its open-access policy, high publication rates in *ACP* correlate least strongly with high citation rates. From linear fits to these publication-citation scatter plots, annual citation count per journal increases by 33 for each additional article published. This annual citation increase per publication is largest for the *Journal of Atmospheric Sciences* at almost 60.

Without filtering for journal, cumulative distributions show that cloud and climate citation peaks for articles written between 1995 and 2005 (Fig. 4a-b). We may expect decreasing citation rates for more recent years, as the time since publication shortens, but the decreasing citation rates prior to 1995 indicate that older literature has been cited less than more recent literature. In the late 1990s, almost 20% of articles have more than a hundred citations, and only 10% are cited five times or fewer, counter to the highly-skewed nature of most citation distributions in other fields in which many articles are never cited and a very small number are very highly cited (Price, 1965).

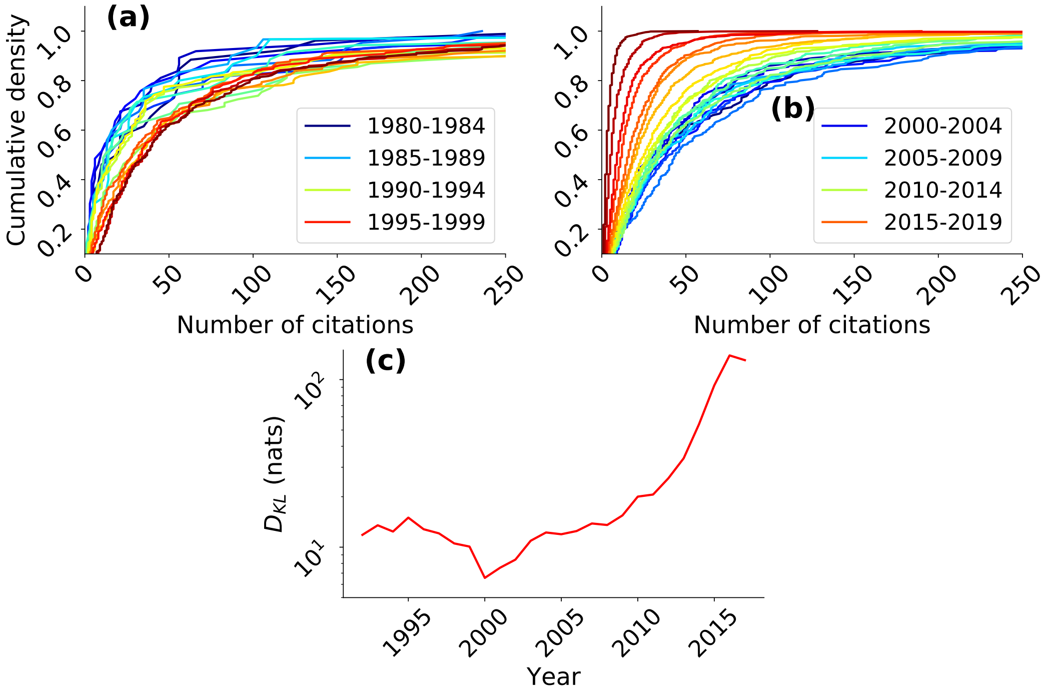

Figure 4. Citation rates in clouds and climate research have slowed relative to publication rate in the most recent decade. Total number of citations per article per year from 1980-1999 (panel a) and from 1999-2019 (panel b). Time series of the Kullback-Leibler divergence between the cumulative density of citations in a given year and the proceeding year (panel c).

To further quantify citation trends, we employ the Kullback-Leibler divergence:

$$D_{KL}(P||Q) = \sum_{x_i} P(x_i)\log_2\left(\frac{P(x_i)}{Q(x_i)}\right) \tag{1}$$

where P and Q are two probability distribution functions of annual citation numbers, x_i. The larger the value of D_{KL}, the more different the distributions P and Q are. Stated more formally, the larger the value of D_{KL}, the more information (in nats when D_{KL} is evaluated with the natural logarithm) would be lost in replacing P by Q. D_{KL} is not symmetric, and we always take the preceding year as the reference distribution (Q). Calculating a 5-year running mean of this pairwise D_{KL} yields relatively stable values of 10 to 15 until 2007; thereafter, D_{KL} increases monotonically up to a most recent value of 140 (Fig. 4c). These increasing D_{KL} values indicate that citation growth in recent years is not maintaining pace with publication growth.

Decreased readability could contribute to this seeming drop in readership in recent years. The assessment of atmospheric science literature in Geerts (1999) found an increasing number of journals, pages per journal, and words per page in articles published between 1965 and 1995. While we reproduce this increasing trend in article length, page number per article, as a crude metric of readability, has been growing far less rapidly than publication or journal numbers, at an average of only a page per decade since 1970 (Fig. 5a). Geerts (1999) also proposed that lagging US federal funding and the plateauing number of PhD students could slow these publication rates in coming years. We find instead that, while US funding has converged or even dropped in recent years, the number of US students in geosciences has continued to grow (Fig. 5b). Equivalent data are not as read-

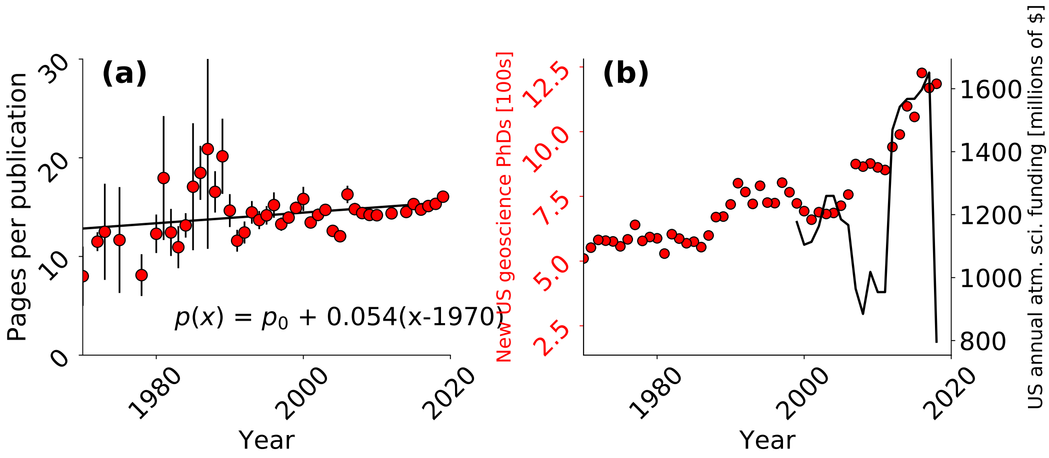

Figure 5. Articles have become only slightly less readable with page number as a metric, while potential readership has doubled since 2000 with graduating doctorates as a metric. Average number of pages per publication per year including standard error and a linear fit in black (panel a). US doctoral degrees awarded in geosciences, including both atmospheric and oceanic sciences, and US federal funding awarded to atmospheric sciences projects (panel b) with data for the last panel come from the US National Center for Science and Engineering Statistics Survey of Earned Doctorates and Survey of Federal Funds for Research and Development (see Sec. 1.2).

ily available for other regions, however, and these US trends may not be representative globally.

3 The Role of Clouds in Radiation, Circulation, and Precipitation

To understand how research effort has been devoted across the climatic impacts of clouds, we next decompose the publication trends from Section 2, using the **Impacts** queries from Table 1. Publications concerning radiation have been most numerous in the past three decades (Fig. 6a), averaging approximately 50 per year in the 1990s, more than twice that many in the 2000s, and more than five times that many in the 2010s. In word clouds of the titles from these publications on clouds and radiation (Sec. 1.2), studies of aerosol and using simulations are dominant, and the most commonly studied regions are the Arctic and the tropics (Fig. 6b). Feedback quantification and the surface energy budget also emerge as common topics.

Publications concerning precipitation have increased most rapidly among the three impacts, doubling every 7 years since 1970 (Fig. 6a). In the associated title word clouds, model simulation and satellite observation appear about equally, and the Arctic and the tropics emerge again as important areas of study, as does China (Fig. 6c). Influence of aerosol and convection or convective parameterization also appear prominently. Perhaps because of the computational demand associated with scale separation, growth has been slowest for publications on clouds and circulation with a doubling time of 10 years. Even more than cloud-radiation studies, these are heavily simulation-oriented (Fig. 6d). Aerosol, feedbacks, and surface interactions are again frequent topics, but less so than for clouds and radiation.

We can further decompose these trends into topics covered in monograph chapters (Fig. 7, **Chapters** queries in Tab. 1). Cloud microphysics is the most active subtheme with almost 400 publications in 2019 alone (Fig. 7 inset). Its growth was particularly rapid in the early to mid-1990s, increasing seven-fold over six years. Precipitation measurement has also been a very active area of research since the mid-1990s with 160 publications in 2019 alone. Slow and steady progress over the past three decades has characterized the optimization of dynamical cores and extension of ground-based measurements. Research output on cloud-circulation coupling has also been more gradual, again possibly because of the computational challenge to simulate all involved scales. Most dramatic is recent growth in machine learning work. While almost no studies used machine learning in 2012, there were 200 publications in 2019 alone.

4 Methodology in Clouds and Climate

4.1 Techniques

The word clouds in Fig. 6b-d reveal some of the most common techniques used in clouds and climate research. We can organize these techniques around pillars of modelling, in-situ or ground-based measurement, and satellite retrieval, and we next search the clouds and climate abstract text in the **Techniques** queries of Tab. 1 (Sec. 1.2).

With this classifications of abstracts, almost 70% mention some kind of modelling (Fig. 8a). Coarse-resolution global climate models (GCMs) still dominate these modelling abstracts (45.8%), although high-resolution or cloud-resolving models account for another 10% each (Fig. 8b). Radiative transfer calculations make up another 18%, while less than 10% of abstracts are attributed to each of large-eddy simulation, trajectory analyses, and idealized or parcel simulations.

But modelling alone characterizes only 22% of abstracts; it is more often used in combination with satellite, flight, reanalysis, or ground-based data (Fig. 8c). Similarly, almost 40% of studies use remote sensing products, but only 6% use remote sensing products exclusively. Ground-based measurements appear in 25% of abstracts, and flight measurements, given their more limited duration, appear in 8%. Laboratory studies are the most infrequent at only 3%, although these abstracts are also hardest to classify with their varied methods and keywords. While models and satellite data provide good spatiotemporal coverage and robust statistics for studies, we conclude that clouds and climate research is fundamentally driven by a synergy of different techniques.

4.2 Authorship

Along with publication trends and techniques, we can also examine the personal aspect of how clouds and climate research is done. Do certain authors always work with the same coauthors? Do two or three “author clusters” characterize certain subfields? We use anonymous Author IDs from the **Impacts** and **Chapters** searches to generate author networks. Circular networks visualize the increase in authorship complexity over time dramatically; in such networks, nodes represent authors, edges represent co-authorship, and node color indicates degree, or coauthor number in our case.

We give an illustrative example of coauthorship evolution in the circular network of Fig. 9, taken from the **Impacts** query on clouds and radiation. In 1989, this network contained only 21 nodes, or authors, with an average degree, or coauthor number, of 1.52. A decade later in 1999, the same network contained 294 authors with an average coauthor number of 3.74. Author number increases by more than an order of magnitude for the other **Impacts** queries in the decade between 1989 and 1999 also (not shown). Authorship in the most recent decade is too complex to present in this layout, but by 2019, author numbers involved in clouds and climate research had increased more than 70-fold since 1989.

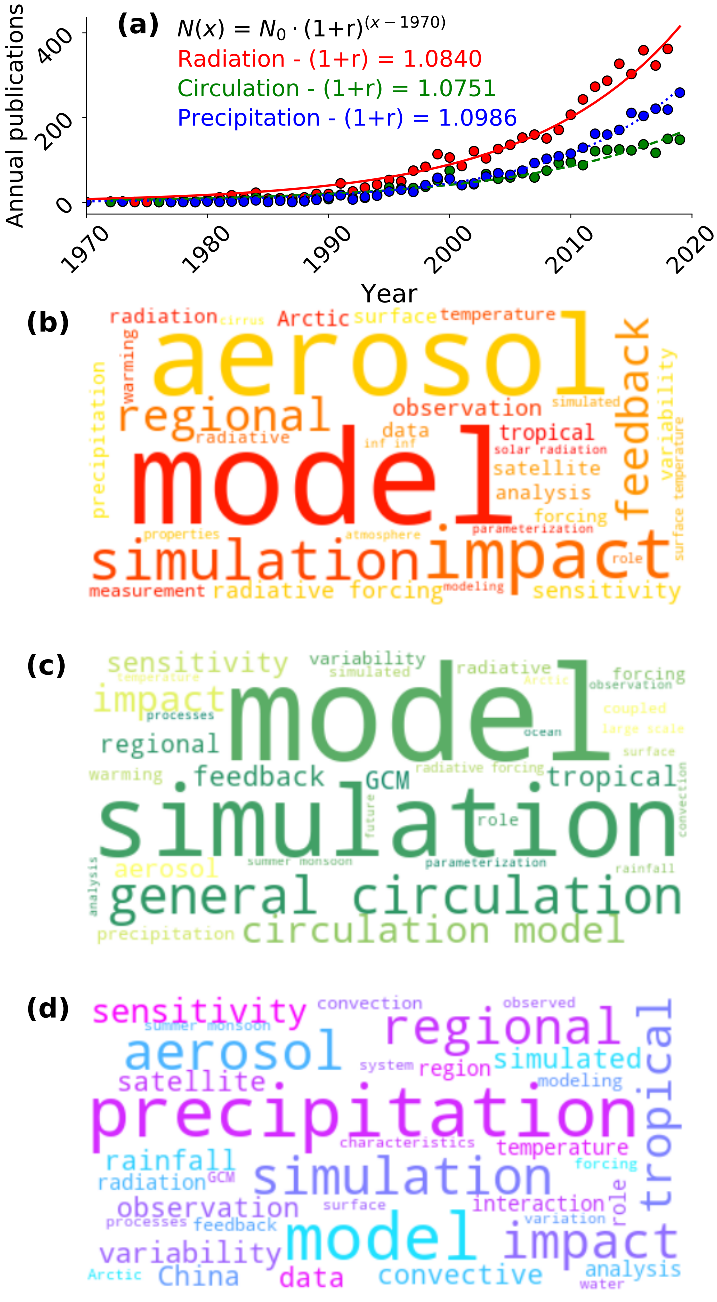

Figure 6. While publication on clouds and radiation have been most numerous, those on clouds and precipitation have increased most rapidly. Publication trends decomposed into cloud impacts on radiation, circulation, and precipitation, as well as their exponential fits, based on the Scopus **Impacts** queries (panel a). Word clouds generated from the article titles for cloud impacts on radiation (panel b), circulation (panel c), and precipitation (panel d). Font size is in proportion to the frequency of the word.

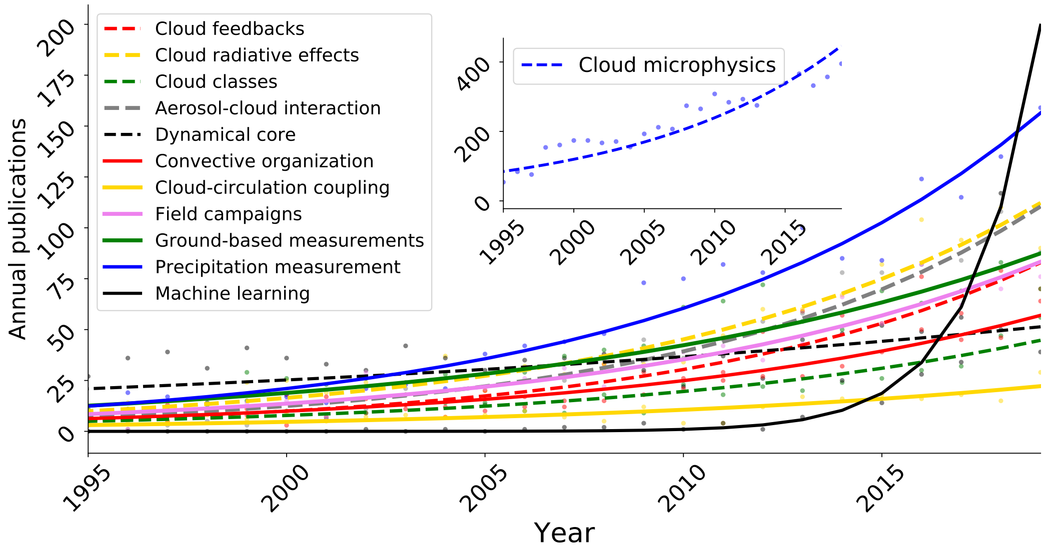

Figure 7. Recent growth in machine learning publication has been dramatic. Cloud microphysics and precipitation measurement are also very active areas of research. Publication trends decomposed into chapter topics, as well as their exponential fits, based on the Scopus **Chapters** queries. The inset y-axis for microphysics publications is two times larger than the primary y-axis.

We can describe networks not only with their degree, but also with their density. Network density refers to the degree normalized by the maximum possible number of edges, or the absolute coauthor number relative to its possible maximum value. Looking again at Fig. 9, while absolute co-authorship increased, relative co-authorship, as quantified by density, decreased six-fold between 1989 and 1999 for work on clouds and radiation. For work on clouds and precipitation, the drop in relative co-authorship is even more dramatic at 30-fold between 1989 and 1999. In the two decades from 1999 to 2019, relative co-authorship dropped another two- to three-fold . The takeaway is that while those in clouds and climate research are collaborating more in absolute terms, they are also directly involved in less and less of the total body of research produced by the community.

Increases in absolute co-authorship may suggest increased objectivity in our results, but as noted above, densification of author networks can actually have an opposite echo chamber effect in which author clusters reinforce their own hypotheses. This kind of clustering cannot be seen in a circular network, so we present authorship networks for the **Chapters** queries also in a spring layout (Fig. 10). A spring layout is generated by a force-directed algorithm in which edges are interpreted as springs governed by Hooke's Law, and nodes are arranged to minimize energy in the network.

These force-directed graphs for authorship tend to have a few studies with big groups of authors and many with smaller groups (Fig. 10). Ideally there would be some overlap and not too much distance between the "hub" studies with large coauthorship and the "rim" studies with limited coauthorship. Work on cloud-radiative effects exhibits such a structure in which collaborations of different size are fairly well mixed, as does work on aerosol-cloud interaction and precipitation measurement, although to a lesser degree. In contrast, work on dynamical cores is characterized by a few studies with large authorship and many smaller-scale efforts that do not overlap at all; however, such separation may be due to model-specific development in this case. We also see the cooperative na-

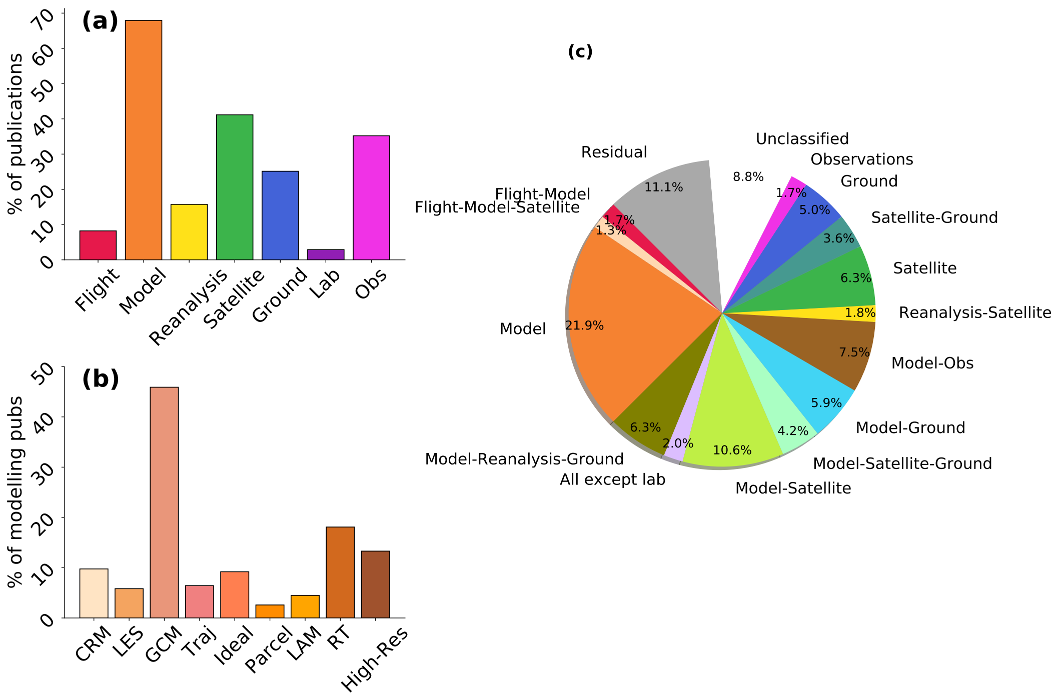

Figure 8. About 40% of studies use a combination of modelling and/or satellite data, but a great diversity of techniques are used to study clouds and climate. Percentage of techniques mentioned in **Clouds and Climate** abstracts, identified in the **Techniques** query (panel a). **Obs** designates unspecified observations, and studies on emissions inventories are grouped into the **Reanalysis** category. Additional filtering of modelling publications into model type (panel b). The following acronyms are used to label: Cloud-resolving model (CRM), global climate model (GCM), trajectories (Traj), idealized or RCE setups (Ideal), limited-area or regional modeling (LAM), radiative transfer calculations (RT), and high-resolution simulation (High-Res). Overall distribution of techniques or technique combinations (panel c). The generic Observations (Obs) classification is removed if another observational classification, e.g. satellite or ground-based was made.

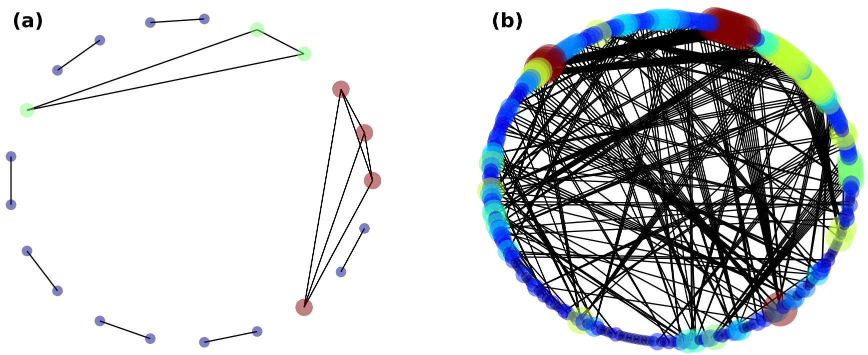

Figure 9. Over a decade, author number increases by a factor of 14 and the degree of co-authorship more than doubles. Evolution of author networks between 1989 (panel a) and 1999 (panel b) for the themes of clouds, climate, and radiation. The higher the degree of the nodes, the larger and darker (more red in the color version) they are.

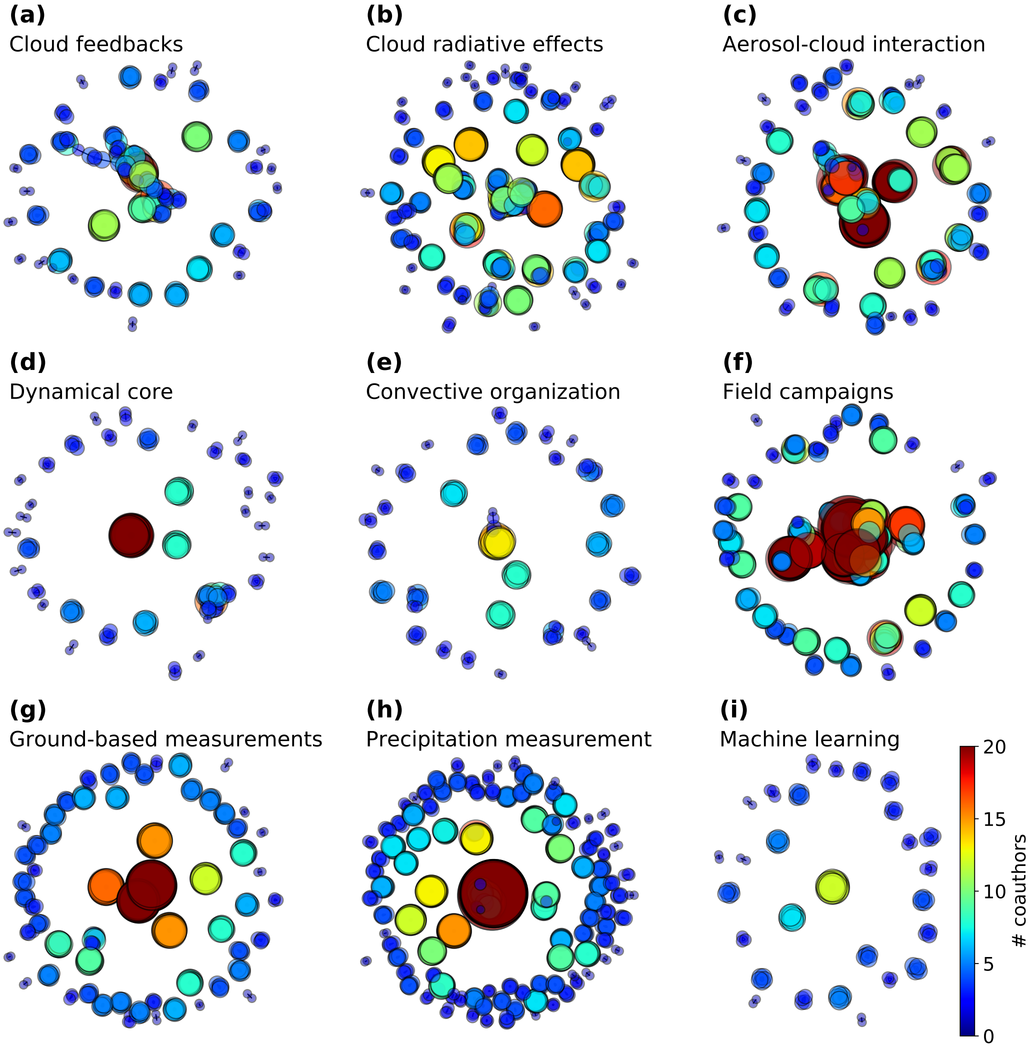

Figure 10. Work across cloud and climate research is typically characterized by a "hub" of studies with large coauthorship and a "rim" of smaller collaborations. Sample author networks for 2015 for several of the chapter themes listed in Tab. 1.

ture of field campaigns with many more "hub" studies than the other subfields, as well as the nascence of machine learning techniques with no large collaborations yet. Such visualizations give us an idea of where author clusters are becoming isolated and where additional community-level work or cross-institutional collaboration could be beneficial.

5 Summary and Outlook

In this introductory chapter, we have used basic SciSci analyses to analyze meta-trends in clouds and climate research. We have also motivated the layout of the overall monograph into impacts of clouds on radiation, circulation, and precipitation. While the largest number of studies exist on clouds and radiation, recent growth in precipitation studies has been most rapid. Output on precipitation measurement is outpaced only by that on cloud microphysics and the recent boom in machine learning. Cloud-circulation coupling remains the least explored impact, given the large range of involved spatiotemporal scales; feasibility of large-domain storm-resolving simulations should bring new insight to this area. We also present three important takeaways from our analyses:

1. **As publications and interest in clouds and climate accelerate, the need to organize and reproduce existing work is becoming more important.** Publication on clouds and climate are doubling every 8 years with almost 2 articles published per day over the last decade. The number of different journals that publish this work is also increasing with a doubling time of 11 years. At the same time, cumulative distributions and distance metrics of citation indicate that readership in recent years may not be keeping pace with this burgeoning literature. As author networks expand, we are less likely to have been involved in or even heard of the efforts from others in our subfield. In the face of these trends, we need to make special effort at the individual or team level to process this information, for example using news aggregators, automatic search alerts, or journal clubs (Landhuis, 2016). Regular meetings of smaller subfield communities are also crucial to promote idea exchange and prevent too much "author clustering".
2. **A synergy of techniques are fueling this growth, and we should continue to take advantage of complementary tools.** Only a third of abstracts on clouds and climate research can be classified as using exclusively satellite data, ground-based measurement, *or* modelling. Instead, it is the combination of simulation and observation that drives the majority of research. We find that almost half of simulation work is still done with coarse-resolution global climate models, although high-resolution models account for another quarter. Increasing computational power means that cloud- and storm-resolving models with kilometer- or hectometer-scale resolutions will become increasingly important in the next few years (Stevens et al., 2020). A new generation of satellite measurements will also become available over the next years, for example with the upcoming launch of EarthCARE or the outcomes of the NASA Aerosol, Cloud, Convection, and Precipitation assessment (Illingworth et al., 2015; Gettelman et al., 2021). We emphasize the role that a synergy of models and measurements have played over the past five decades and its continued importance going forward.
3. **Small-scale cloud processes constitute the highest publication numbers, suggesting their intractability. Orienting the recent surge in machine learning toward these problems could afford progress.** Publications in microphysics, precipitation measurement, and aerosol-cloud interaction have been the most numerous over the past ten years. These research areas represent the smallest scales involved in cloud formation and evolution: droplet activation and coalescence and ice nucleation and aggregation. In the past five years, work in machine learning on clouds and climate has also exploded from 7 publications in 2012 to 198 in 2019. How machine learning techniques can assist in the persistent small-scale problems of cloud physics is worthy of exploration. Such efforts are already underway, for example in using Bayesian inference for rain microphysics or neural networks to represent aerosol effective radii and refractive indices (e.g., Morrison et al., 2019; Llerena et al., 2018). Neural networks have also been used to represent aspects of raindrop formation by droplet collision-coalescence and could be an even more promising route for ice microphysics, characterized by closed-form solutions or property tables rather than the coupled differential equations involved in the stochastic collection equation (e.g., Gettelman et al., 2020; Seifert & Rasp, 2020).

While these SciSci analyses have been straightforward, they still give insight into which topics and tools are receiving the most attention. We suggest that, with its synthesis of scientometrics and sociology, science of science can also complement traditional review to help us understand discrepancies and biases in how clouds and climate research is done. For example, publication data mining through tools like PaperHunter (`http://`

paperhunter.net), Connected Papers (https://www.connectedpapers.com), or ScholarSight (http://scholarsight.org) could be used to pinpoint contradictory studies or quantify the range in uncertain parameters across many studies. With more comprehensive citation trees or text mining for technique identification, we would lose track of past and originating ideas less readily, avoid repeating certain studies unnecessarily, or reproduce other studies more intentionally. Such meta-level analyses would allow us to answer questions like, Are our hypotheses biased by authorship clusters? *or* Are we providing sufficient funding to the intersection of subfields? Given the socioeconomic costs of climate change, clouds and climate research is done in a uniquely urgent environment, and we posit that science of science could help to optimize how this work is done in the decisive years ahead.

Acknowledgments

SS acknowledges funding from DFG Project 02.056750161 (TropiC) and from the German Ministry of Education and Research (BMBF) and FONA: Research for Sustainable Development (www.fona.de) under grant 01LK1509A. CH acknowledges funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme under grant agreement No 714062 (ERC Starting Grant "C2Phase"). Both authors thanks two anonymous reviewers for their feedback that has improved the manuscript. SS thanks Aiko Voigt for feedback on an early version of this work and Philipp Gasch for discussions about this topic. Information about databases and search terms are given in Methods and Table 1, and the results of these searches is also directly available at https://doi.org/10.5281/zenodo.5524758. Codes to reproduce figures are provided at https://github.com/sylviasullivan/CloudLit.

References

Boucher, O., Randall, D., Artaxo, P., Bretherton, C., Feingold, G., Forster, P., … Zhang, X. (2013). Clouds and aerosols. in: *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press.*

Douville, H., Raghavan, K., Renwick, J., Allan, R. P., Arias, P. A., Barlow, M., … Zolina, O. (2021). Water cycle changes. in: *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* [in press]. *Cambridge University Press.*

Forster, P., Storelvmo, T., Armour, K., Collins, W., Dufresne, J.-L., Frame, D., … Zhang, H. (2021). The earth's energy budget, climate feedbacks, and climate sensitivity. in: *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* [in press]. *Cambridge University Press.*

Fortunato, S., Bergstrom, C. T., Börner, K., Evans, J. A., Helbing, D., Milojević, S., … A.-L-Barab'asi (2018). Science of science. *Science*, *359*(6379). doi: 10.1126/science.aao0185

Foster, J. G., Rzhetsky, A., & Evans, J. A. (2015). Tradition and innovation in scientists' research strategies. *Am. Soc. Rev.*, *80*(5). doi: 10.1177/0003122415601618

Geerts, B. (1999). Trends in atmospheric science journals: A reader's perspective. *B. Amer. Meteorol. Soc.*, *80*, 639–652.

Gettelman, A., Carmichael, G. R., Feingold, G., Silva, A. M. D., & van den Heever, S. C. (2021). Confronting future models with future satellite observations of clouds and aerosols. *Bull. Amer. Meteorol. Soc.*, *102*(8), E1557–E1562. doi: 10.1175/BAMS-D-21-0029.1

Gettelman, A., Chen, J., & Gagne, D. J. (2020). Machine learning rain formation and impacts on clouds, precipitation, and radiation..

Illingworth, A. J., Barker, H. W., Beljaars, A., Ceccaldi, M., Chepfer, H., Clerbaux, N., . . . van Zadelhoff, G.-J. (2015). The EarthCARE satellite: The next step forward in global measurements of clouds, aerosols, precipitation, and radiation. *Bull. Amer. Meteor. Soc.*, *96*(8), 1311-1332.

Landhuis, E. (2016). Scientific literature: Information overload. *Nature*, *535*, 457-–458.

Llerena, C., Müller, D., Adams, R., Davey, N., & Sun, Y. (2018). Estimation of microphysical parameters of atmospheric pollution using machine learning. In (pp. 579–588). Springer.

Milojević, S. (2015). Quantifying the cognitive extent of science. *J. Inform.*(9), 962–973. doi: 110.1016/j.joi.2015.10.005

Morrison, H., van Lier-Walqui, M., Kumjian, M. R., & Prat, O. P. (2019). A Bayesian approach for statistical-physical bulk parameterization microphysics. Part I: Scheme description. *J. Clim.*, *77*, 1019-1041. doi: 10.1175/JCLI-D-19-0070.1

Price, D. J. S. (1963). *Little science, big science.* Columbia Univ. Press.

Price, D. J. S. (1965). Networks of scientific papers. *Science*, *149*(3638), 510–515. doi: 10.1126/science.149.3683.510

Seifert, A., & Rasp, S. (2020). Potential and limitations of machine learning for modeling warm-rain cloud microphysical processes. *J. Adv. Model. Earth Sys.*, *12*(12). doi: 10.1029/2020MS002301

Stevens, B., Acquistaoace, C., Hansen, A., Heinze, R., Klinger, C., Klocke, D., . . . Zängl, G. (2020). The added value of large-eddy and storm-resolving models for simulating clouds and precipitation. *J. Meteorol. Soc. Jap.*, *98*, 395-435. doi: 10.2151/jmsj.2020-021

Webb, M. J., Andrews, T., Bodas-Salcedo, A., Bony, S., Bretherton, C. S., Chadwick, R., . . . Watanabe, M. (2017). The Cloud Feedback Model Intercomparison Project (CFMIP) contribution to CMIP6. *Geosci. Model Dev.*, *10*(1), 359–384.