



Connecting scientific advances and patented technologies: The role of open access scientific publishing in clean-technology innovation

Benedict **Probst**, Andreas **Kontoleon**, Laura Diaz **Anadon**

C-EENRG Working Papers, 2021-4

March 2021

Please cite this paper as:

B. Probst, A. Kontoleon, L. D. Anadon, 2021, 'Connecting scientific advances and patented technologies: The role of open access scientific publishing in clean-technology innovation'. C-EENRG Working Papers, 2021-4. pp.1-28. Cambridge Centre for Environment, Energy and Natural Resource Governance, University of Cambridge.

The Cambridge Centre for Environment, Energy and Natural Resource Governance (C-EENRG) was established in 2014 within the Department of Land Economy in order to conduct integrative research on the governance of environmental transitions, such as the climate-driven transition from a carbon-intensive inefficient energy matrix to a decarbonised and efficient one, or the water/population-driven transformation of food production systems. The role of law and governance in such transitions is of particular importance. C-EENRG approaches law as a technology to bring, guide and/or manage environment-driven societal transformations. Whereas other research groups within the University of Cambridge cover questions relating to the environment, energy and natural resources from a natural science, engineering, conservation and policy perspective, until 2014 there was no centre concentrating on the law and governance aspects of environmental transitions. C-EENRG was established to fill this gap in the policy research cycle, in line with University's strategic research initiatives on energy, conservation, global food security and public policy.

The **C-EENRG Working Papers** provide a platform to convey research aiming to support the effective governance of such environmental transitions. The series hosts research across disciplines, including law, economics, policy, and modelling, and welcomes research on a wide range of environmental topics across the food-energy-water-land nexus.

SCIENTIFIC COMMITTEE

Professor Laura Diaz Anadon	<i>Climate policy, economics and transitions</i>
Professor Andreas Kontoleon	<i>Economics and biodiversity</i>
Dr Shaun Larcom	<i>Economics and biodiversity</i>
Dr Emma Lees	<i>Law and governance</i>
Dr Jean-François Mercure	<i>Modelling and transitions</i>
Dr Pablo Salas	<i>Modelling and transitions</i>
Professor Jorge E. Viñuales	<i>Law, governance and transitions</i>

Send your enquiries regarding the C-EENRG Working Papers to the editorial team: **Paul Lohmann** pml44@cam.ac.uk and **Dr Tong Xu** tx224@cam.ac.uk

GUIDELINES FOR AUTHORS

Submit your manuscript in *.odt, *.doc, *.docx, or *.rtf format, together with a *.pdf version. It should contain title, abstract, keywords, and contact details for the lead author. Email your file to the editor. The manuscript will be processed and sent to two reviewers. The editor will process the reviewers' feedback and request the scientific committee to adopt a final decision about the publication of the manuscript

DISCLAIMER

The opinions expressed in this working paper do not necessarily reflect the position of C-EENRG, the Department of Land Economy, or the University of Cambridge as a whole.

C-EENRG, 2021, <http://www.ceenrg.landecon.cam.ac.uk/>

Benedict Probst

Cambridge Centre for Energy, Environment and Natural Resources Governance (C-EENRG),
Department of Land Economy, University of Cambridge, UK

Andreas Kontoleon

Cambridge Centre for Energy, Environment and Natural Resources Governance (C-EENRG),
Department of Land Economy, University of Cambridge, UK

Laura Diaz Anadon

Cambridge Centre for Energy, Environment and Natural Resources Governance (C-EENRG),
Department of Land Economy, University of Cambridge, UK

Contact:

Benedict Probst

Silver St., CB3 9EU, Darwin College, Cambridge, UK

Cambridge, UK

bsp26@cam.ac.uk

Connecting scientific advances and patented technologies: The role of open access scientific publishing in clean-technology innovation

Benedict Probst, Andreas Kontoleon, Laura Diaz Anadon

ABSTRACT

Getting to a net-zero economy requires faster development and diffusion of novel clean energy technologies. Understanding how the characteristics of scientific publications, such as the degree of accessibility, impact the development of technological innovations is critical. Yet, empirically establishing the links between scientific research and technological application (often studied using patents) is a highly under-researched area due to its complexity. Combining data science methods, machine learning, and econometrics, we develop and apply a novel approach to assess the determinants of the probability with which scientific discoveries (codified in scientific articles) are used in four major types of clean energy patents (i.e., batteries, biofuel, solar power, and wind power). We use data from more than 100,000 scientific articles and over 600,000 patents from these four technology categories for the period 2005 to 2018. Based on this data we first evaluate whether the diffusion time between scientific publishing and patented technologies has changed. We then investigate the effects of the characteristics of scientific articles on the probability that an article will be used in a patent as measured by citations in the non-patent literature. We also evaluate the effects of the characteristics of scientific articles that have been used by patents on the actual influence of the patent itself on other patents. We find that the average lag between the publication of the scientific article and the citation of the patent has sharply decreased from on average 5-6 years in 2005 to less than 2 years in 2013. We also find that the most important predictor of whether a scientific article is used in a patent is whether it can be assessed without a paywall. In addition, we find that scientific articles that are cited more often by other scientific articles also lead to more influential patents. Our results indicate that the current scientific publication market structure – characterised by an oligopoly of publishers primarily relying on subscriptions from universities – is associated with a small but adverse effect on clean-tech innovation. Our results have profound implications for the transition to a carbon net-zero economy and could also have wider implications in other sectors (such as pharmaceuticals and biotechnology).

Keywords: Open Access, Patenting, Clean Technology, Technological Innovation

1. Introduction

Achieving net-zero carbon emissions by the second half of this century as envisaged by the Paris Agreement (Viñuales, 2015) requires – among other things – faster development and diffusion of novel clean energy technologies (Bosetti *et al.*, 2009). Although technology innovation is a complex process that goes beyond research and development and includes diffusion (Anadon *et al.*, 2016) and obsolescence (Wilson *et al.*, 2012), the flow of knowledge from scientific discoveries – typically (but not always) codified in scientific publications – into patents is an important part of it (Dechezleprêtre *et al.*, 2011).

One of the factors that could shape the likelihood of scientific publications influencing the development of technologies is the extent to which inventors globally have knowledge about and access to those publications. As the predominant model of scientific publishing continues to rely on a subscription-based system, inventors without access to major journals paid by their universities, laboratories or companies may have difficulties accessing scientific publications (Tennant *et al.*, 2016). While the number of open access (OA) journals, publications in paywalled journals published as OA (so-called ‘gold OA’), and articles made available in repositories (so-called ‘green OA’) has increased substantially over the last decade, the most highly-cited academic journals maintain a subscription-based model of publishing (Piwowar *et al.*, 2018).

Over the past five years, several universities and other public research institutions ranging from the University of California to the Max Planck Institute have voiced an emerging resistance to the subscription-based system adopted by most academic journals. Some of the reasons for this resistance are perceptions of unfair pricing and the assertion that the results of publicly-funded science should be available to the public. In addition, several funders (such as all UK’s Research Councils and the EU’s research funding programs) already require at least ‘accepted’ versions of academic articles published in subscription-based journals to be freely available in repositories under the rationale that the knowledge funded by the public sector should be easily and publicly available (Else, 2018). The most radical OA initiative unveiled in 2018 is dubbed ‘Plan-S’, which will require scientists receiving funding from certain European funders controlling EUR 7.6 billion to make their research directly available upon publication from 2021 onward, at least in a repository (Else, 2018).

Previous research studying the impact of research available OA across different academic disciplines – such as natural and social sciences – relying on regression analyses without directly addressing selection bias have found that, *ceteris paribus*, OA is associated with a positive impact of various types of subsequent citations in other academic publications (Guédon, 2004). More robust empirical strategies relying on randomisation and other statistical approaches trying to address selection bias – the possibility that scientists send their best articles to subscription-based journals – have resulted in lower estimates of these effects on the probability of subsequent citation in academic publications (Davis and Fromerth, 2007; Davis, 2011; McCabe and Snyder, 2014). Thus, overall, the literature indicates that OA publication of research – either

through repositories, gold OA or OA journals – is associated with higher academic impacts (Piwowar *et al.*, 2018). Yet, the relationship between the degree of accessibility of a scientific publication and its subsequent impact on technological applications – in other words, commercial and wider economic impact – proxied by patents is, however, unclear.

We, therefore, contribute to filling this gap in the literature by proceeding as follows. First, we develop a new approach to trace the incorporation of scientific articles into patented technologies. Similar to scientific papers, patents are required to cite knowledge that is relevant or has provided a stepping stone for the invention seeking patent protection. Most of these citations are to other patents, but patents can also cite non-patent literature that was relevant to the invention. We rely on these non-patent citations and develop a novel empirical methodology to establish the aforementioned link between patents citing academic publications. Second, we investigate to what extent the diffusion time of academic knowledge and patented technologies have changed over time. Third, reliant on the developed approach, we are able to then uniquely empirically assess the determinants of the probability that scientific discoveries (proxied by scientific articles) are used (i.e. cited) in patents. We do so by using data from 102,301 articles (of which 12,876 are open access) and ~600,000 patents covering four clean energy technologies between 2005 and 2018. We select four important clean energy technologies, namely batteries, biofuel, solar power, and wind power), as they have experienced major advances over the past 15 years. Lastly, we investigate how the characteristics of scientific articles that patents rely on affect follow-on innovation (measured by citations from other patents in the five years after the patent’s publication). In other words, we delve deeper into assessing how more influential patents are impacted by the characteristics of the scientific articles they relied on (such as metrics typically used to assess their quality and degree of accessibility).

Our main hypothesis explored in the empirical analysis described above is based on the literature on frictions in information gathering (Wilson, 2012; McCabe and Snyder, 2014) and stipulates that, all else being equal, academic publications that are harder or more expensive to access or less visible (because it may take more time to find them in open repositories than on journal websites), will have a lower probability of being used in patents by researchers, inventors, and entrepreneurs seeking to develop technologies to address problems or meet market needs, thereby reducing the probability that they will use these scientific insights. The extent to which this hypothesis is true and some innovators are unable to access or have more difficulties accessing scientific publications could matter in many areas of science and technology that ‘may affect the probability and speed with which we tackle various societal challenges, from mitigating climate change (which is the challenge explored here), to developing a COVID-19 vaccine or seeds that are more resistant to droughts in the developing world.

Shedding light on this issue is of particular importance for implementing the aggressive de-carbonization agendas adopted by countries all around the world. Research has shown that delaying clean-technology innovation can have a big impact on climate change mitigation, perhaps even bigger than even delayed climate policy, defined as putting in place strict CO₂ emission limits (Luderer *et al.*, 2012). Hence, understanding

the extent to which the current model of scientific publishing may be affecting the probability of academic findings being used in clean energy technology patents is critical.

The remainder of the paper is organised as follows. Section 2 details our approach to link scientific insights to eventually patented technologies and describes the related data and empirical strategy. Section 3 presents results from our descriptive and statistical analysis on the diffusion time and the determinants that lead to scientific papers being cited by patents. Section 4 then shows how characteristics of scientific studies that patents rely on influence follow-on innovation (measured by citations from other patents in the five years after publication). Section 5 analyses the scientific articles behind the most influential patents. Section 6 concludes and discusses the policy implications of these findings.

1.1. Linking Scientific Publications and Patented Technologies

We use developments in natural language processing, machine learning, and econometrics to develop a novel method for tracking the diffusion of scientific knowledge into patented technologies. Existing work analysing the diffusion of scientific or technological advances primarily focuses on publication to publication citations (Davis, 2011) or patent to patent citations (Acemoglu, Akcigit and Kerr, 2016). This is mainly because these links are much more difficult to establish and extract. Similar to scientific papers, patents are required to cite knowledge that has been useful in the innovation process. While patents mostly cite other patents, they also cite scientific publications. Utilizing this information (of scientific publication-patent links) would allow for improved mapping of the diffusion of scientific knowledge into technological discoveries and applications. To establish this link between scientific papers and patents, we proceed in eight steps, described in detail in the SI and briefly laid out below. While the approach in this paper is used to study clean energy technologies, it can be used in any area of science and technology where scientific publishing and patenting plays a role – covering most industrial sectors, such as medical and agricultural technologies, digital communication, and energy (EPO, 2019).

The first and second step in our approach entails the selection of relevant patents via International Patent Classification (IPC)¹ codes and keywords, which represent technological classes (Figure 1). While a patent can be assigned several IPC classes, it generally has one main IPC class. Adding keywords has been shown to remove substantial false positives (i.e., patents that show up in the data belonging to that particular technological class, such as digital cameras when searching for solar PV patents) and is commonly used across the literature (Huenteler *et al.*, 2016). While using topic modelling instead of IPC codes would be more fine-grained (Suominen, Toivanen and Seppänen,

¹ The International Patent Classification came into force through the Strasbourg Agreement in 1971. It established an hierarchical system for the classification of patents, allowing the segregation of patents into distinct, fine-grained technological areas (WIPO, 2019).

2017), it is too computational-intensive for very large datasets and might therefore rather be used on a subset of patents that have already been identified via IPC codes.

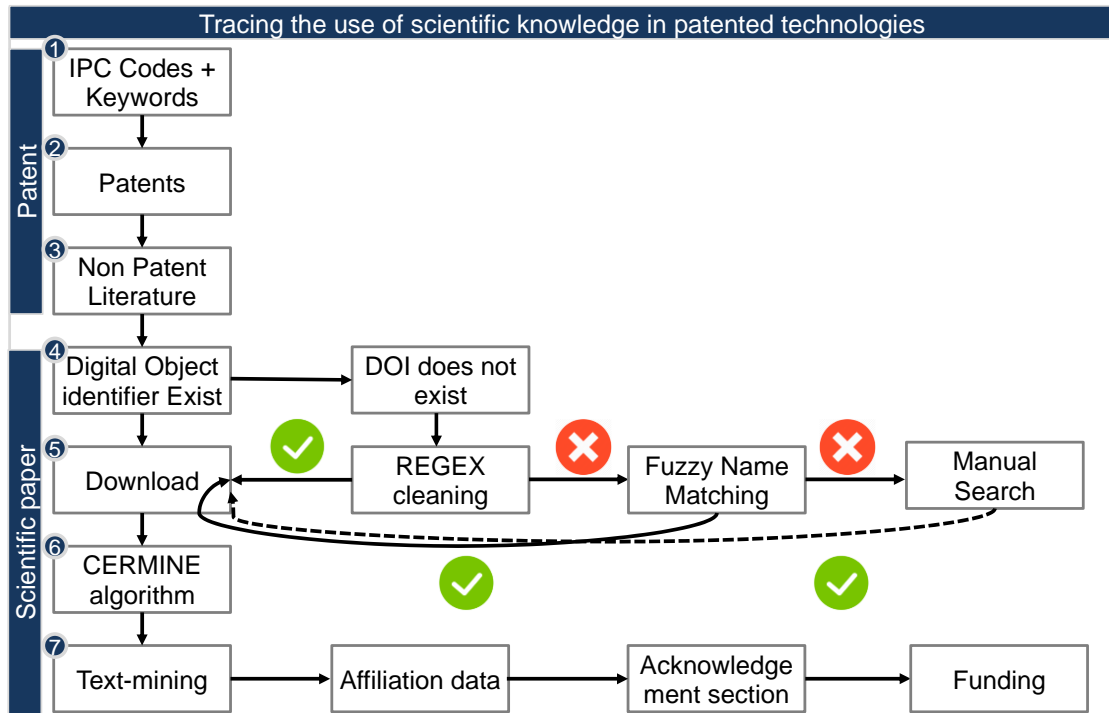


Figure 1: Patent-scientific publication linking approach developed in this paper. REGEX cleaning refers to a filtering process using Regular Expressions (REGEX), which is a sequence of characters used to find a specific pattern within the data (e.g., in many patents the digital object identifier (DOI) is not provided, but may be hidden within the description of the patent. Hence one could look for filtering ‘doi’ to find see whether it is provided somewhere in the patent). Source: author

In a third step, each patent’s reference list to non-patent literature is extracted. Non-patent literature that can be cited by patents includes scientific publications (code: ‘s’), chemical abstracts, and patents by the Japanese authority. For our purpose, we only extract scientific publications that have been cited by the patent.

In a fourth step, the digital object identifier (DOI) of the scientific article referenced in the patent is extracted – if the DOI exists in the patent database. The DOI is a persistent identifier that allows for the unique identification of digital objects, such as academic citations. It is maintained by the International Organisation of Standardisation (ISO) and it covers more than 85 million publications from more than 10,000 organisations. The DOI is crucial to extract metadata through the API (Application Programming Interface) of online databases. These include CrossRef, which is an open-source, non-profit database that stores information on more than 100 million

publications, which include author names, journal, publication year, and cited references. However, it commonly does not contain – in its metadata– information on the authors’ affiliation, funding, or the acknowledgement section. These can only be found in the actual text of the publication, not in the metadata. To understand the geographical characteristics of university-firm knowledge flows, however, this approach had to also reveal the affiliation of the scientists, since it reveals important geographic information (e.g., whether it is an international collaboration) and the acknowledgement section often shows which organisation funded the research. The latter information is particularly important for evaluating the effectiveness of R&D funding (which goes beyond the scope of this paper).

The digital object identifier is extracted. Yet, this step is not without challenges. In contrast to citing other patents, there are no clear rules on how to cite scientific publications in patent applications. This leads to many unstructured entries. We use several approaches in that case. First, we devise several regular expression operators, which allow us to search for text-bits such as “DOI:” in unstructured data and allow us to extract relevant entries if they exist. If this does not work, we use a large database of journal articles from the Web of Science (identified via key-word searches) and match them via fuzzy-matching procedures. This approach does not match ‘exactly’ but allows for a certain leeway in naming conventions or reference styles. Our fuzzy matching relies on the Levenshtein ratio, which is a distance measure between our original text-string and the text-string that is to be matched. If that approach does not work, we turn to manual approaches, which entails searching the publication by hand. This very work-intensive approach was only done in exceptional circumstances.

Fifth, once we have isolated the DOI – either directly or indirectly – we download the publication via several R-procedures. We combine different packages, such as R CrossRef, which checks each paywalled publisher whether the article exists for a given DOI. This can be done on thousands of articles at once and is only constrained by download speed and request limits, which differ across publishers. Yet, with the procedure a large number of articles can be downloaded within minutes. Articles commonly are either downloaded in PDF or XML format. Whereas PDF refers to portable document format and is widely used in publishing, XML is the acronym for Extensible Markup Language. The latter is machine-readable and structured in a hierarchical tree-format. Some publishers only have a PDF file in their repositories, which is true particularly for older research (e.g., scans of papers).

As the ultimate goal is to have an XML format, which is more readily usable for text mining, we use the CERMINE algorithm to convert PDF files to XML files (Tkaczyk *et al.*, 2018). The CERMINE algorithm, developed at the Interdisciplinary Centre for Mathematical and Computational Modelling at the University of Warsaw in Poland, employs an unsupervised machine-learning approach to convert PDF files to XML.

There are a number of alternatives to the CERMINE algorithm, but we make use of it for several reasons. First, it is written in open source code, which allows researchers to understand the mechanics of the algorithm instead of blindly relying on a black-box algorithm. Second, the algorithm performs well when benchmarked against other algorithms (such as ParsCit or GROBID) in terms of precision and recall (Tkaczyk *et*

al., 2018). Third, it offers a greater range of functionality when it comes to extracting metadata, such as author affiliation data (critical to understand the geographic configuration of research) and acknowledgement sections (important to extract funding information). As this information is not commonly available through metadata databases, such as CrossRef, the only viable current way is to use algorithms to extract it directly from the article.

The CERMINE algorithm works as illustrated in Figure 2. First, the CERMINE algorithm is fed a PDF file. It then proceeds with a layout analysis, where it creates a hierarchical structure of the document (while preserving the entire text). This first step is composed of three sub-steps, which is 1) character extraction (size and position on the page), 2) page segmentation (geometric hierarchical structure with zones, which contains lines of words, which in turn are composed of characters) and 3) reading order determination (which usually is top to bottom). After the layout analysis, the second step entails content classification, which contains two sub-steps: 1) initial zone classification and 2) metadata zone classification. During this first sub-step, each zone is classified into metadata (authors' name, affiliation, etc.), references, body (the main text of the article), or other (acknowledgements and conflict of interest). In the second sub-step, the initial coarse classification is further refined. For instance, metadata zones are classed into more granular metadata classes, such as title, editor, and type.

These classifiers use support vector machines, which are supervised machine-learning algorithms. These algorithms are trained on datasets that already contain the correct labels (e.g., title and author affiliation, among others) and then commonly rely on a non-probabilistic binary linear classifier to classify previously unseen datasets (or more specifically, scientific articles, as in our application). The support vector machines rely on the LibSVM library (Chang and Lin, 2011), a widely used support vector machine package to solve optimisation problems. In a third step, the extracted metadata is then analysed by 1) metadata zone classification (assigning more granular metadata classes to metadata zones) and extracting metadata from labelled zones. In a fourth step, the bibliography is extracted, which contains references and the related metadata (author, title, issue, etc). Again, the zones are then classified into more granular parts.

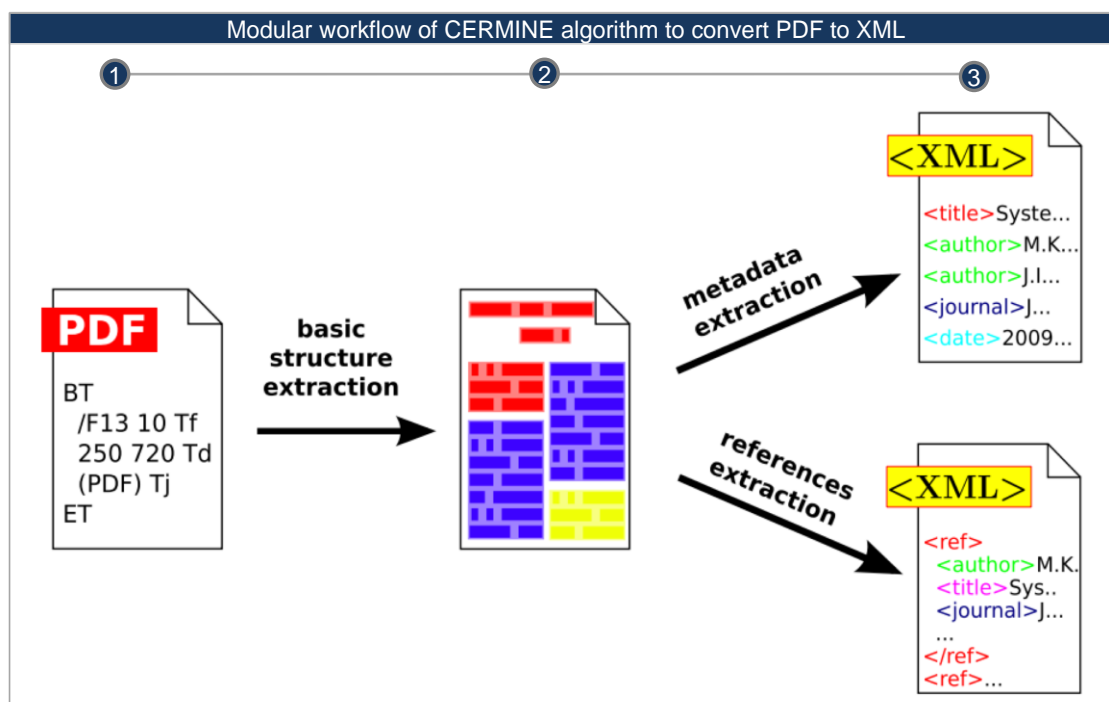


Figure 2: Modular workflow of CERMINE algorithm to convert PDF to XML. Source: (Tkaczyk et al., 2018)

The final output of the CERMINE algorithm is an NLM JATS (National Library of Medicine Journal Article Tag Suite), which is a rich composition of XML elements and attributes. JATS stores a host of structured metadata, such as title, author, etc, as well as the full text, bibliography, which applies the National Information Standards Organisation (NISO) Z39.96-2012 standard and can therefore be widely used by libraries and research organisations.

Seventh, after the JATS is extracted it is imported into R. In R a range of different R-packages is used to extract the relevant metadata from the JATS output, such as the XML package. The XML package relies on the hierarchical structure of the JATS files to extract metadata information for each article and returns a dataset with the specific metadata.

Lastly, we use the keywords developed in Popp (2016) and our own keywords to identify and download relevant scientific articles in the technological fields this study covers, but can be extended to any technological field in which scientific articles and patenting play a role.

2. Data and Methods

2.1. Data

Our final dataset contains 102,301 scientific articles downloaded from the Web of Science (of which 12,876 are open access), which we link to ~600,000 patents in four clean energy technologies between 2005 and 2018. Patent data is from the European Patent Office database (PATSTAT) Autumn Edition 2018. It is important to note that PATSTAT does not only hold European patents but includes 100 million patents from 90 major patent issuing authorities from around the world and covering the mid 19th century to 2019 (PATSTAT, 2016).

Before detailing our empirical analysis in the next sections, a few notes on the descriptive statistics of our dataset of publications and patents on those four energy technologies (Table 1 for full details).

Descriptive Statistics								
<i>Full Dataset (2005-2013)</i>								
<i>Publications</i>	#n	Cited by Patent	Closed Access	Full Open Access (Gold)	Open Access Repository (Green)	Open Access Journal	Average number authors	Average citation
All	102,301	3.5%	86.7%	4.8%	3.5%	5.0%	4.8	43.3
By Technology								
Batteries (Lithium)	20,251	5.2%	92.8%	2.8%	1.5%	3.0%	4.98	54.1
Biofuels	30,379	3.2%	82.2%	6.8%	6.1%	4.9%	4.46	40.0
Solar PV	44,980	3.4%	89.4%	2.6%	2.6%	5.4%	5.27	42.3
Wind	6,691	0.50%	81.3%	4.3%	4.4%	10.0%	3.38	32.0

Note: #n is the number of articles published in each category. Open Access Publications are the average of publications that were open access, Cited by Patent indicates whether the scientific publication was cited by a patent, Open Access Journal whether the article was published in an open access journal or not (Note that for a publication do to be open access it does not have to be published in open access journals). Number of authors represents the number of authors on each scientific article. Source = Web of Science and PATSTAT Autumn 2018 Edition. Percentage may not sum to 100% due to rounding

Table 1: Descriptive Statistics of full dataset

Of the 102,301 scientific articles across the four technologies included in our analysis, only 3.5% eventually get cited by patents. This seemingly low number may be explained

by several factors: First, there is a skewed distribution of the impact of scientific research on technological progress, with some articles having an outsized influence. For instance, it is estimated that around one third of scientific publications in the natural sciences and engineering account for more than 80% of citations received from other scientific publications (Larivière, Gingras and Archambault, 2009). Second, knowledge accumulation is commonly described as cumulative, contributing to the ‘knowledge stock’ (Acemoglu, Akecigit and Kerr, 2016). Hence, many publications may contribute to an article that is eventually cited by a patent. Third, while patent examiners do a thorough search of prior knowledge a particular patent relies on, they may miss some academic articles that may have been important to the development of the patent itself. Lastly, another point worth highlighting is that that articles published open access – either through an OA journal, gold, or repository – have increased across all technologies, though for the observed time period they have remained under 20% of all published outputs (Figure 3).

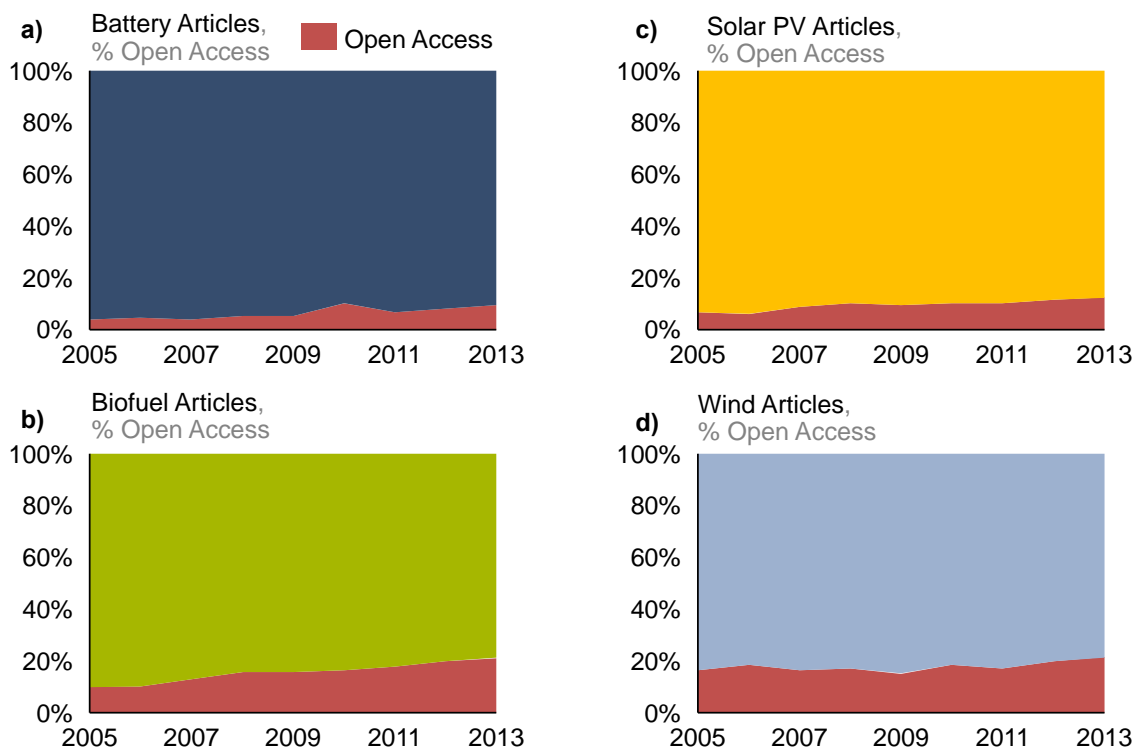


Figure 3: Open access publication trends across four clean technology fields between 2005-2013 (all articles identified via keywords, regardless of whether cited by patent or not). Only scientific publications before 2014 are included because we weigh each patent by a five-year citation period to account for strongly different ‘values’ of patents. We cut the patents in 2018 because the last two years tend to be cut-off due to delays in the patenting process. Specifically, scientific articles related to a) batteries, b) biofuels, c)

solar PV articles, and d) wind articles from Web of Science (WoS) based on keyword search developed by Popp (2016) and our own keywords.

Specifically, scientific articles related to a) batteries, b) biofuels, c) Solar PV articles, and d) wind articles from Web of Science (WoS) based on keyword search developed by Popp (2016).

2.2. Methods

Our statistical analysis determining the drivers of patents citing academic publications consists of two steps. The first regression tests what determines the likelihood of a scientific publication to be used in a patent. For this analysis we use scientific and patent data from 2005-2013. In a second regression, we only use those scientific papers that eventually got cited by a patent to investigate which characteristics of these scientific studies influence follow-on innovation (measured by citations to the patent citing the academic publication by other patents in the five years after publication). Hence, in the second step (Section 4) we delve deeper into the determinants of high-impact patents. For instance, we pose the question of whether academic success also predicts the impact on technological application, proxied through the number of patents that rely on the research. For this second regression we use scientific and patent data from 2005-2018. The longer time frame used for this analysis stems from the need to weigh ‘patent influence’ by the citations received from other patents in the first five years of publications. Studies show that citation-weighted patent counts are more strongly correlated with a patent’s influence and market value than unweighted patent counts (Hall, Jaffe and Trajtenberg, 2001).

In the first regression, our outcome variable is binary (0 = not cited by patent, 1 = cited by patent) and therefore use a logistic regression (King and Zeng, 2003). The raw coefficients have to be transformed into average marginal effects, as effect sizes for logistic regressions depend on other independent variables in the model.

We run the logistic regression in equation (1):

$$p_i = \alpha_0 + \beta\phi_i + \zeta_i + \omega_i + \partial_i + \varrho_i + \varepsilon_i \quad (1)$$

where p_i is a dummy that indicates whether scientific publication i got cited by at least one patent; ϕ_i represents the characteristics of that scientific publication (number of authors, technological field, and year of publication). ζ_i indicates whether the paper is gold OA, ω_i whether it can be accessed in an OA repository, and ∂_i is whether it was

published in an open-access journal (i.e., in which all articles are OA by default). ϱ_i is a technology dummy. α is the constant and ε is the error term.

To address concerns over the possibility that authors select their best articles into some of the top paywalled journals or pay for their free publication in gold open access or into some of the most prestigious journals, such as Nature and Science, which are not free OA (also known as selection bias), we use various propensity score matching techniques based on topic modelling via the Latent Dirichlet Algorithm (LDA is applied to the text of the scientific publications in our data; a detailed description of the methods can be found in the SI). This allows us to achieve a greater homogeneity in co-variables (i.e., number of authors, year of publication, and subject matter) across the treated and untreated group of publications (i.e., OA and non-OA). In other words, it allows us to only compare articles that are similar in terms of article quality, content, and other relevant confounders (see Supplementary Note 1 for a detailed discussion).

Generally, for large samples (as in our case), the choice of the matching algorithm is less important than for smaller samples, as “asymptotically all PSM estimators should yield the same results, because with growing sample size they all become closer to comparing only exact matches” (Caliendo and Kopenig, 2005; p.11). We use the nearest neighbour matching algorithm for our matching procedure because the algorithm has been shown (in Monte Carlo Simulations) to perform well in comparison to other matching algorithms (Austin, 2014). (See Supplementary Note 2 for a detailed discussion).

Yet, the regression in (1) does not reveal whether patents relying on OA scientific publications are more or less influential than those that rely on non-OA articles in terms of subsequently patented inventions. We now turn to the question of how the most influential patents (identified using the common practice of the number of forward citations) are affected by the characteristics of the scientific articles that these patents have cited. This is important because research has shown that the influence of patents is heavily skewed (Harhoff *et al.*, 1999), a phenomenon that has been referred to as the ‘tail in the tail’ (i.e., even within the top-5% of influential patents, only a fraction of these patents get the most citations from other patents). We investigate the relationship between scientific publications cited by patents in the four energy technologies and the influence of the patents measured by patent citations five years after the publication of the patent.

Given the count nature of our dependent variable, we run a negative binomial regression model. We estimate the regression in equation (2):

$$\psi_i = \alpha_0 + \beta\Phi_i + \zeta_i + \omega_i + \varrho_i + \varepsilon_i \quad (2)$$

where ψ_i is the weighted count of patent citations that a particular patent has received in the first five years following its publication, whereas Φ_i represents the characteristics of a given paper (number of authors). ζ_i indicates whether the paper is full OA and ω_i

whether it can be accessed in an OA repository. q_i is the technology dummy (batteries, biofuel, solar power, and wind power). α is the constant and ε is the error term.

We find that the number of citations and number of authors of the scientific article that a patent relies upon are good predictors of the 5-year forward citations the patent receives from other patents. This is consistent with research trying to understand the factors behind the quality of scientific publications, which found that article citations and number of authors are good predictors of research quality—more citations are often used as a proxy for more useful research, in part because bigger teams might have more resources as well as a more diverse knowledge base and connections (Ductor, 2015). In our analysis, we find that these predictors do not only lead to a higher likelihood of a scientific publication being cited by a patent in the non-patent literature, but they are also associated with more impactful patents (both in terms of the number of patents relying on those articles and the number of and follow-on citations patents receive from other patents (so-called forward citations)).

3. Results

3.1. Overall decrease in diffusion time across all papers

The general diffusion lag from public R&D to its eventual application can be described in three steps: 1) Public R&D funding that 2) generates a new scientific article that is eventually 3) cited by a patent. Existing evidence on the lag between 1) and 2) for clean energy technologies indicate that it takes between 2-10 years from initial R&D funding to publication of the scientific article (Popp, 2016). In this section, we investigate the lag between steps 2-3.

Our first central finding relying on the approach developed in Section 2 is that the average lag between the publication of the scientific article and the citation of the patent has sharply decreased from on average 5-6 years in 2005 to less than 2 years in 2013 (Figure 4). This is shorter than previous research has found using multinomial lag models, and indicates that this research might have overestimated the actual lag-length that it takes from research to application. Our findings on the diffusion of scientific knowledge to patented technologies are important to inform other research that needs to make assumptions about the time-lag between research and diffusion (Popp, Hascic and Medhi, 2011). Generally, it is encouraging to see, that research appears to find its way quicker into application than previously thought.

The lag length does not differ significantly between technologies. For all four technologies, the lag between scientific research and eventual application has decreased from between 4-7 years to slightly under two years in 2013 (except wind, where the data for 2005-2008 is limited). This indicates that the diffusion of knowledge has accelerated between 2005-2013.

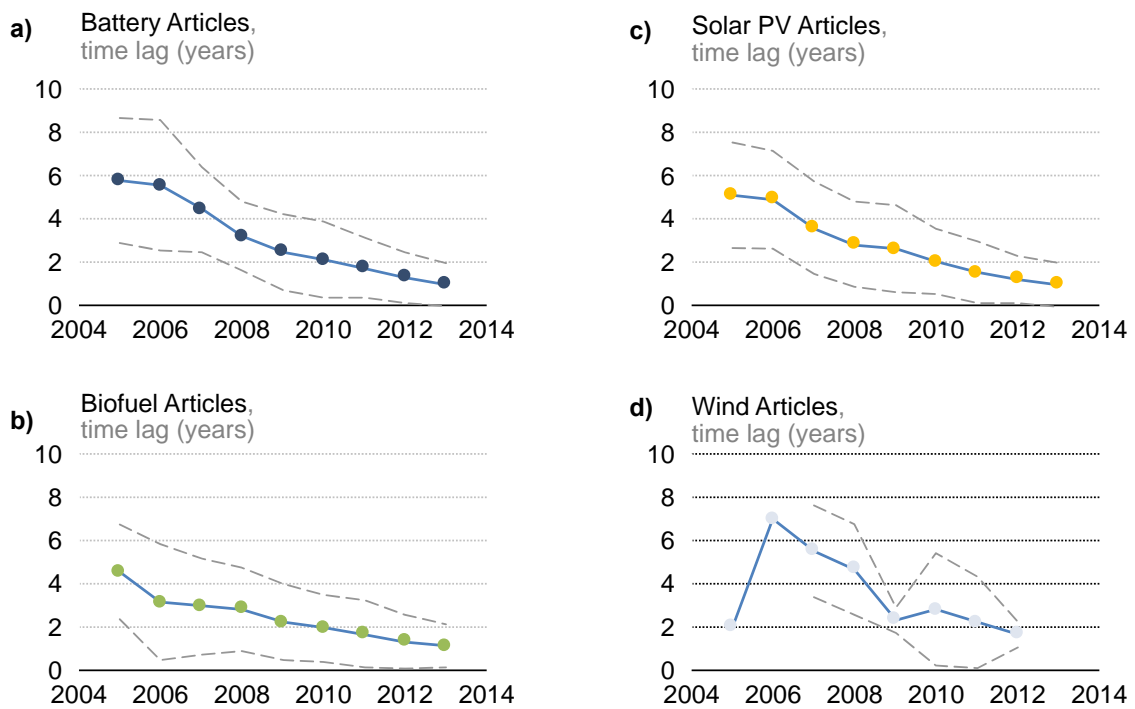


Figure Error! No text of specified style in document.: Trends in the diffusion lag between scientific articles and patent citation. Blue line indicates average years and dotted grey lines are one standard deviation error bars for: (a) batteries, (b) biofuels, (c) solar PV and (d) wind.

As this trend is observed across all technologies, this may indicate that larger forces – that touch all four technologies – are at work. These likely include increased competition and technological maturation (Hoppmann *et al.*, 2013; Huenteler *et al.*, 2016), the ascent of fast and widespread communication technologies (Greenstein, 2010) and declining research productivity (hence, needing more research to maintain the same level of output) (Bloom *et al.*, 2017). It is also important to note that public and private R&D investment in renewable energy more than doubled from 2005-2013 from USD 4.0 to 9.2 billion, likely also decreasing the time it takes from scientific articles to patent citation. The increased R&D investments were also driven by greater attention to climate change, notably with the award of the Nobel Peace Prize to the IPCC in 2007.

The heightened awareness of climate change also led to several public support schemes for renewable energy, such as the feed-in tariff scheme for photovoltaics in Germany starting in the early 2000s creating the first mass-market in Europe (Nemet, 2010). As the market size for clean technology inventions was growing, associated inventions were also becoming more valuable. Research has also repeatedly demonstrated that more valuable patents will be granted substantially earlier than less valuable

patents, likely because patent holders attempt to fast-track more valuable patents through the request for accelerated examination (Harhoff and Wagner, 2009). Similarly, research also finds that patents with higher generality – hence, patents that are cited by a more diverse range of subsequent patents – are also more valuable and are examined quicker.

Also, the European Patent Office has also reduced the time it takes for examination due to various measures to speed up the patenting process. For instance, at the beginning of the 2000s, the *Bringing Examination and Search Together* (BEST) project was rolled out at the EPO to speed up examination times while keeping the quality. The BEST project led to a sizeable reduction in the time it took examiners to complete their search reports, which forms the basis for the grant decision of the patent. The number of months it took to complete a search report fell from 10 months in 2003 to 7.8 months in 2004 and has since stabilised at around 6 months, a reduction by 40% over time. Taken these trends together, it stands to reason that both the lag structure and the decline in examination time between technologies should be similar.

3.2. Increased patent-paper citation likelihood associated with OA journals

After having discussed the average decline of the diffusion time, we now turn towards the main part of our empirical analysis. In this section we analyse not the speed, but the likelihood that a scientific research article eventually gets used in a patent.

Our results show that academic citations, the number of authors, and OA are all positively and significantly correlated with the likelihood of an academic article being used in a patent, although the magnitude of these effects is small (see SI for full regression tables). These factors have also been shown to play a role in increasing the likelihood of citation in analyses of paper-to-paper citation data (Tahamtan, Safipour Afshar and Ahamdzadeh, 2016). We show now (utilizing scientific publication-patent citation links) that these factors are also important in the development of four clean energy technologies.

The effect of different types of OA in particular demonstrates that articles published in open access repositories and journals drive the entire effect, whereas gold open access does not have a significant effect at the 5% level (Figure 5). The difference between open access repositories and open access journals could stem from differences in publication timing, as open access journals immediately publish articles whereas sharing articles placed in repositories is subject to various forms of restrictions during the embargo period (normally lasting 6-24 months (Elsevier, 2019)). The lack of an effect on patent citations of gold open access publications could suggest that researchers and inventors who do not have access to paywalled journals may not even go on the websites of journals where the majority of articles is paywalled (even if some articles on the website

are golden access). For these researchers or inventors, gold OA may therefore not be different in practice from closed access.

We find that open access articles are between 0.8% (repositories) and 1.6% (OA journals) more likely than paywalled articles to be cited by a patent. While the effect might not seem high, it is important to note that one single breakthrough patent could have a substantial influence on the pace and depth of decarbonising the world economy. This is, for example, the whole model of energy R&D funding agencies like ARPA-E (Goldstein *et al.*, 2020) and of the VC model, in general, rely on low-probability high-impact innovations. To verify the results, we run a range of sensitivity and robustness checks which are included in the SI, which include testing different matching strategies. These checks produced similar results corroborating the robustness of our analysis.

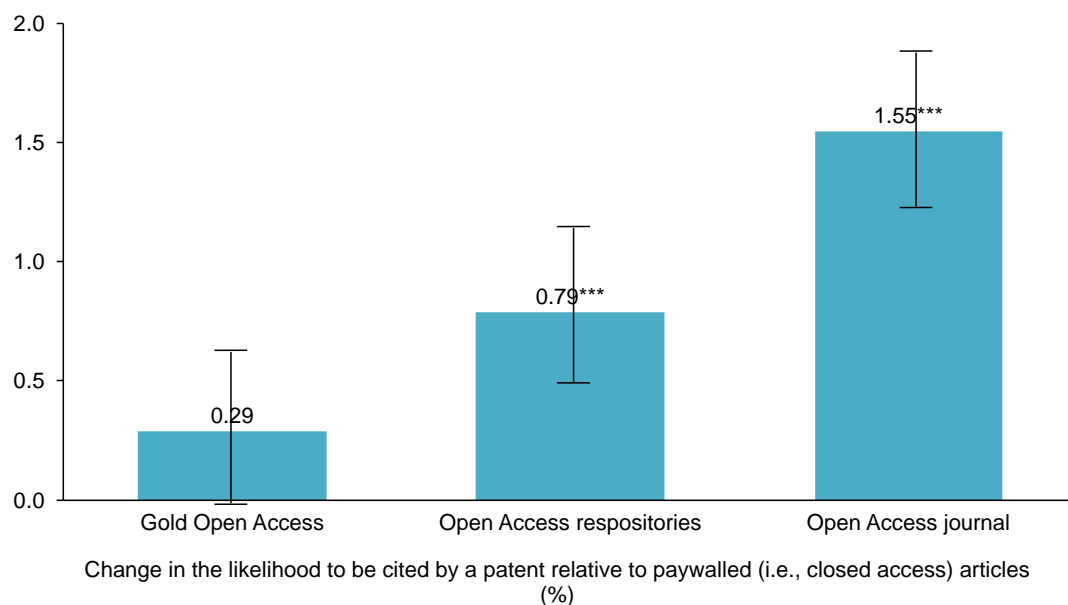


Figure 5: Paper-patent citation likelihood by open-access type relative to the baseline (closed & paywalled journal) controlling for the quality of the paper and other relevant covariates. To make the impact comparable, we match the scientific papers on content through topic modelling and other important covariates. Gold open access refers to

journal articles that are published in a subscription-based journal, but authors pay to have the piece published without a paywall. *** refer to statistical significance at 1%.

3.3. More influential scientific articles lead to more influential patents

The previous analysis indicates that controlling for the research area, number of authors and year of publication, scientific articles in the four energy technologies covered under different types of OA access are more likely to be cited in patents. Yet, our analysis did not reveal whether patents relying on OA scientific publications are more or less influential than those that rely on non-OA articles in terms of subsequently patented inventions. We now turn to the question of how the most influential patents (identified using the common practice of the number of forward citations) are affected by the characteristics of the scientific articles that these patents have cited.

We find that the number of citations and number of authors of the scientific article that a patent relies upon are good predictors of the 5-year forward citations the patent receives from other patents. This is consistent with research trying to understand the factors behind the quality of scientific publications, which found that article citations and number of authors are good predictors of research quality—more citations are often used as a proxy for more useful research, in part because bigger teams might have more resources as well as a more diverse knowledge base and connections (Ductor, 2015). In our analysis, we find that these predictors do not only lead to a higher likelihood of a scientific publication being cited by a patent in the non-patent literature, but they are also associated with more impactful patents (both in terms of the number of patents relying on those articles and the number of and follow-on citations patents receive from other patents (so-called forward citations)).

We find that patents citing OA articles, controlling for patent characteristics, are also associated with higher forward patent citations compared to the ‘closed category’ (Table 2). But the magnitude of the effect between the different OA categories and patents of higher value (as proxied by forward citations) is now different. Gold open access articles – which have the lowest likelihood of being cited by patents of all OA categories as shown in Figure 5– generate the largest number of follow-on patent citations (of all OA categories). In contrast, patents relying on publications in OA repositories and OA journals – which had the second-highest and highest likelihood of being cited by patents – generate fewer follow-on citations from other patents. One possible explanation for this is that scientists are more likely to pay for Gold OA if they think that the findings could have a sizeable impact beyond academia (since most scientists will have access through their own academic institution even if the publication is paywalled, but inventors outside of the academia may not. This could be tested in future work).

Negative Binomial Regression Full Dataset					
Weighted Number of Patent Citations					
	(1)	(2)	(3)	(4)	(5)
Citations	0.004 *** (0.00014)	0.003*** (0.0001)	0.003*** (0.0001)	0.003*** (0.0001)	0.003*** (0.0001)
N_Authors			0.088*** (0.016)	0.082*** (0.016)	0.082*** (0.016)
Is Open Access				0.613*** (0.141)	
Open Access Disaggregated					
<i>Base: Closed</i>					
OA-Repository					0.665*** (0.211)
Gold open access					0.992*** (0.248)
Open Access Journal					0.099 (0.244)
Year Controls	No	Yes	Yes	Yes	Yes
N	3,573	3,573	3,573	3,573	3,573
AIC	19,932	19,747	19,714	19,698	19,700

Table 2: Main results of patent-impact regression analysing the impact of OA on patent success (includes tech controls)

Overall, the impact of scientific publications on patents is heavily skewed. Figure 6a maps the citation network structure of around 100,000 articles and 600,000 patents, which shows that a large proportion of scientific articles is never cited by patents, but a few scientific articles generate large follow-on citations: 10% of scientific publications generate 44% of citations from patents in the four clean technologies we study (Figure 6b).

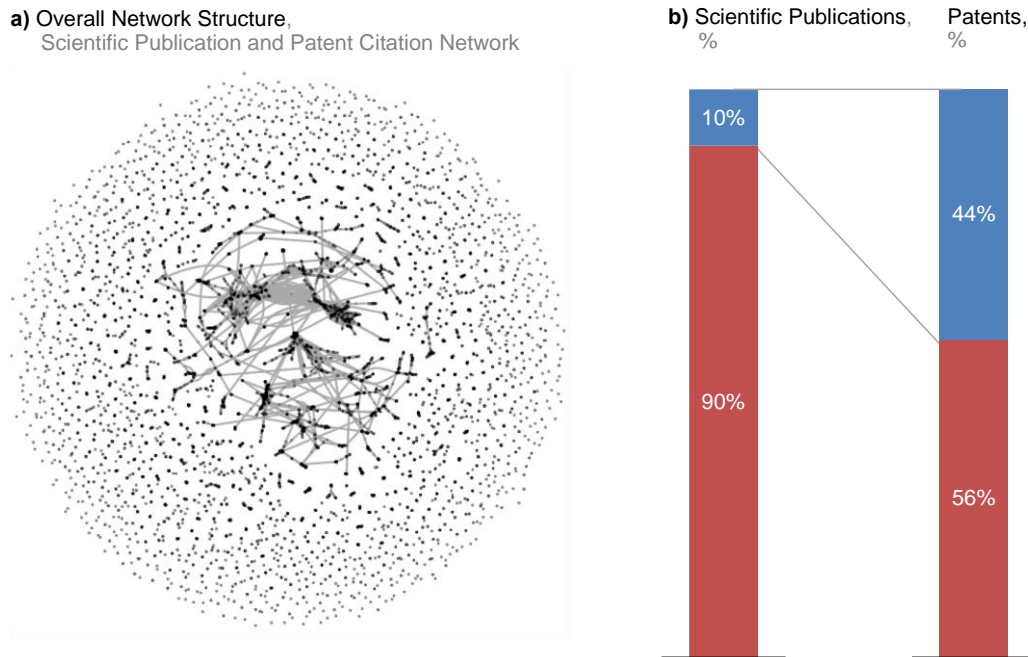


Figure 6: a) Overall network structure of scientific publication and patent citation network 2005-2018; and b) skewness of impact of scientific publications on patent filings for the four clean-energy technologies. Note: Patents in b) are weighted by citations received in the five years after the publication of the patent to account for differences in the value of the four clean-energy technology patents. Nodes in a) are the citations between scientific publications and patents. Shows that very few peer-reviewed academic citations lead to substantial follow-on citations. Colours in 3b only for illustrative purposes (blue indicating top 10% of scientific publications leading to 44% of all patents).

3.4. Most influential patents rely on paywalled articles

Looking at the top-100 cited scientific research articles, particularly the top-tier scientific articles published on solar PV, we see that they appear to have had a substantial impact on patenting and follow-on innovation (Figure 7). For instance, Burschka et al. (2013) Nature publication *Sequential deposition as a route to high-performance perovskite-sensitized solar cells* has so far received 5290 citations from other scientific articles and was cited in 19 patents with 30 follow-on citations on average from other patents over the next five years. Even more impressive, Lee et al (2012) Science publication on *Efficient Hybrid Solar Cells Based on Meso-Superstructured Organometal Halide Perovskites* has received a similar citation count from other scientific articles (6231) and received substantially more patent citations (44 patents with 402 citations on average).

The two examples of papers mentioned above are in the upper right quadrant of Figure 7, displaying publication-patent links with both high patent and high academic impact. While there are some scientific publications that are highly cited both by other

scientific publications and patents that are open-access within this top-100 of publications across the four technology areas, the field continues to be dominated by paywalled publications, particularly in journals such as Nature and Science.

Most academic publications are in the first quadrant, which has low patent and academic impact. Yet, this is only true in comparison to the most impactful papers in the dataset. As papers in the first quadrant have between 0 – 80 times the average citation impact of their field in a specific year (as 1 indicates the average citation impact of all papers in a given field in a given year; see Figure 7).

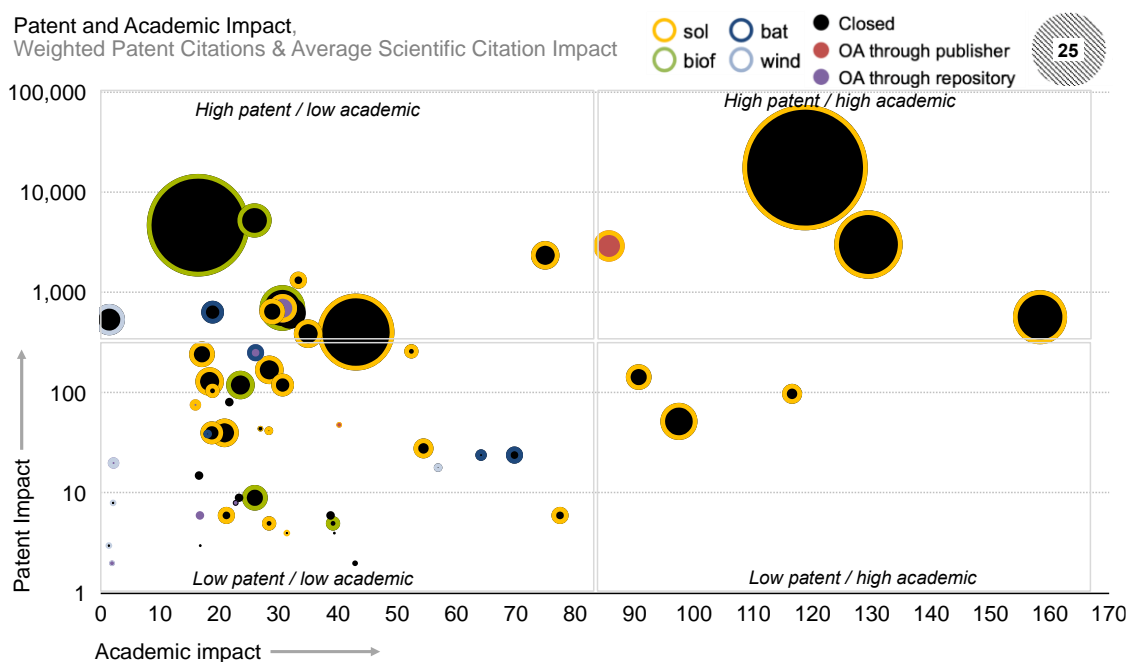


Figure 7: Patent and academic impact for solar PV, li-ion batteries, biofuels, and wind energy technologies for top-100 paper-patent pairs. The category ‘OA through repository’ includes both open access journals as well as open access repositories, which are not shown separately here due to low occurrences. The size of the bubble indicates the number of patents that rely on academic research. Black filling indicates closed publication, red repository, and blue OA via publisher. Yellow lining indicates solar, green biofuel, dark blue batteries and light blue wind. Academic impact measured as average citation impact (ACI), which sets the average number of citations received by a given field of research in a given year to 1. Hence, 10 indicates that a given article received 10 times the citations of the average article in the field in a given year. Patent impact is the number of patents that have cited the paper in the first five years, weighted by the citations that these patents received themselves. ‘Sol’ refers to solar, ‘biof’ to biofuels, ‘bat’ to batteries. Closed are scientific articles that are behind a paywall, OA through publisher are those that are published in an OA journal or have been made OA on a publishers’ website (gold access), which are put into one category here due to low

occurrence. Repository refers to those articles that are behind a paywall, but there is an OA repository that holds an openly accessible version.

4. Implications for Science and Energy Policy

Our results point to possible important implications for science and energy policy. First, it appears that the current scientific publication market structure – characterised by an oligopoly of publishers primarily relying on subscriptions from universities – is associated with a small but adverse effect on cleantech innovation measured by the probability that cleantech patents will rely on the scientific publications. It appears that some inventors find it more difficult to access scientific knowledge for the development of patents when the publications are not available in OA. This could have a detrimental impact on cleantech innovation and subsequent diffusion. Our empirical results underscore findings from theoretical modelling that the availability of climate change mitigation technologies is critical to reaching net-zero carbon emission goals (Luderer *et al.*, 2012). Yet, overall the reduction in diffusion time between scientific research and clean-technology patenting is encouraging.

Second, our research also provides some evidence supporting the notion that academic impact measured by academic citations is, on average, associated with patents that are then more widely cited and relied upon by other newer patents. This can be a proxy for higher possible commercial impact. The finding that more highly-cited academic papers also lead to more highly-cited patents, indicates that publication citations (at least in this area of science and technology) may be a useful factor to consider when assessing the impact of academic researchers. Our results provide evidence against the common critique that citation indices are too short-sighted and do not capture influence beyond the academic system, as we show that academic citations are a good indicator of subsequent patent citations (Costas and Franssen, 2018), while recognizing that the role of citations differ by academic fields and can only ever be part of a complex assessment process.

Lastly, the approach developed in this paper to link scientific publications and technological patenting has wide applicability. Existing approaches to study vertical spillovers are largely limited to case-study evidence and interviews (Agrawal and Henderson, 2002), which is difficult to scale across different technological areas and research institutes. Hence, our approach can be used in any area of science and technology where scientific research and patenting play a role, such as clean energy technologies, telecommunications, chemicals, biotechnology, and pharmaceuticals.

ACKNOWLEDGEMENTS

This work was supported by the European Union's Horizon 2020 research and innovation programme project INNOPATHS [Grant agreement No. 730403], the Department of Land Economy and the School of Humanities and Social Sciences, University of Cambridge and the Heinrich Böll Foundation.

REFERENCES

- Acemoglu, D., Akcigit, U. and Kerr, W. R. (2016) ‘Innovation network’, *Proceedings of the National Academy of Sciences of the United States of America*, 113(41), pp. 11483–11488. doi: 10.1073/pnas.1613559113.
- Anadon, L. D. *et al.* (2016) ‘Making technological innovation work for sustainable development’, *Proceedings of the National Academy of Sciences*, 113(35), p. 201525004. doi: 10.1073/PNAS.1525004113.
- Austin, P. C. (2014) ‘A comparison of 12 algorithms for matching on the propensity score’, *Statistics in Medicine*, 33(6), pp. 1057–1069. doi: 10.1002/sim.6004.
- Blei, D. M. *et al.* (2003) ‘Latent Dirichlet Allocation’, *Journal of Machine Learning Research*, 3, pp. 993–1022. doi: 10.1162/jmlr.2003.3.4-5.993.
- Bloom, N. *et al.* (2017) ‘Are Ideas Getting Harder to Find?’ doi: 10.3386/w23782.
- Bosetti, V. *et al.* (2009) ‘The Role of R&D and Technology Diffusion in Climate Change Mitigation: New Perspectives Using the WITCH Model’, *Energy*, pp. 1–41. doi: 10.1787/227114657270.
- Caliendo, M. and Kopenig, S. (2005) ‘Some Practical Guidance for the Implementation of Propensity Score Matching’, *Journal of Economic Surveys*, 22(1), pp. 31–72. Available at: <http://ftp.iza.org/dp1588.pdf%5Cnpapers3://publication/uuid/B68613C9-4FF0-4002-B086-60831932B7D5>.
- Chan, G. (2015) *Essays on Energy Technology Innovation Policy*. Available at: <http://nrs.harvard.edu/urn-3:HUL.InstRepos:17467190%0AThis>.
- Chang, C.-C. and Lin, C.-J. (2011) ‘LIBSVM’, *ACM Transactions on Intelligent Systems and Technology*, 2(3), pp. 1–27. doi: 10.1145/1961189.1961199.
- Costas, R. and Franssen, T. (2018) ‘Reflections around “the cautionary use” of the h-index: response to Teixeira da Silva and Dobránszki’, *Scientometrics*. Springer Netherlands, 115(2), pp. 1125–1130. doi: 10.1007/s11192-018-2683-0.
- Davis, P. M. (2011) ‘Open access, readership, citations: A randomized controlled trial of scientific journal publishing’, *FASEB Journal*, 25(7), pp. 2129–2134. doi: 10.1096/fj.11-183988.
- Davis, P. M. and Fromerth, M. J. (2007) ‘Does the arXiv lead to higher citations and reduced publisher downloads for mathematics articles?’, *Scientometrics*, 71(2), pp. 203–215. doi: 10.1007/s11192-007-1661-8.
- Dechezleprêtre, A. *et al.* (2011) ‘Invention and transfer of climate change-mitigation technologies: A global analysis’, *Review of Environmental Economics and Policy*, 5(1), pp. 109–130. doi: 10.1093/reep/req023.
- Ductor, L. (2015) ‘Does co-authorship lead to higher academic productivity?’, *Oxford Bulletin of Economics and Statistics*, 77(3), pp. 385–407. doi: 10.1111/obes.12070.
- Else, H. (2018) ‘Radical open-access plan could spell end to journal subscriptions’, *Nature*, 561(7721), pp. 17–18. doi: 10.1038/d41586-018-06178-7.
- Elsevier (2019) *Journal Specific Embargo Periods 2018*. Available at: https://www.elsevier.com/___data/promis_misc/external-embargo-list.pdf.

- EPO (2019) 'Trends in patenting'. Available at: [http://documents.epo.org/projects/babylon/eponet.nsf/0/5B8C2A3F8EFBBB82C1258247003F549D/\\$File/epo_annual_report_2017_infographic_en.pdf](http://documents.epo.org/projects/babylon/eponet.nsf/0/5B8C2A3F8EFBBB82C1258247003F549D/$File/epo_annual_report_2017_infographic_en.pdf).
- Goldstein, A. *et al.* (2020) 'Patenting and business outcomes for cleantech startups funded by the Advanced Research Projects Agency-Energy', *Nature Energy*. Springer US, 5(October). doi: 10.1038/s41560-020-00683-8.
- Greenstein, S. (2010) *Nurturing the Accumulation of Innovations*; *Management*. doi: 10.3386/w15905.
- Guédon, J.-C. (2004) 'The "Green" and "Gold" Roads to Open Access: The Case for Mixing and Matching', *Serials Review*, 30(4), pp. 315–328. doi: 10.1016/j.serrev.2004.09.005.
- Harhoff, D. *et al.* (1999) 'Citation Frequency and the Value of Patented Inventions', *Review of Economics and Statistics*, 81(3), pp. 511–515. doi: 10.1162/003465399558265.
- Harhoff, D. and Wagner, S. (2009) 'The duration of patent examination at the european patent office', *Management Science*, 55(12), pp. 1969–1984. doi: 10.1287/mnsc.1090.1069.
- Hoppmann, J. *et al.* (2013) 'The two faces of market support - How deployment policies affect technological exploration and exploitation in the solar photovoltaic industry', *Research Policy*. Elsevier B.V., 42(4), pp. 989–1003. doi: 10.1016/j.respol.2013.01.002.
- Huenteler, J. *et al.* (2016) 'Technology life-cycles in the energy sector - Technological characteristics and the role of deployment for innovation', *Technological Forecasting and Social Change*, 104, pp. 102–121. doi: 10.1016/j.techfore.2015.09.022.
- King, G. and Zeng, L. (2003) 'Logistic regression in rare events data', *Journal of Statistical Software*, 8, pp. 137–163.
- Larivière, V., Gingras, Y. and Archambault, É. (2009) 'Brief communication the decline in the concentration of citations, 1900-2007', *Journal of the American Society for Information Science and Technology*, 60(4), pp. 858–862. doi: 10.1002/asi.21011.
- Luderer, G. *et al.* (2012) 'The economics of decarbonizing the energy system-results and insights from the RECIPE model intercomparison', *Climatic Change*, 114(1), pp. 9–37. doi: 10.1007/s10584-011-0105-x.
- McCabe, M. J. and Snyder, C. M. (2014) 'Identifying the Effect of Open Access on Citations Using a Panel of Science Journals', *Economic Inquiry*, 52(4), pp. 1284–1300. doi: 10.1111/ecin.12064.
- Nemet, G. F. (2010) 'Robust incentives and the design of a climate change governance regime', *Energy Policy*. Elsevier, 38(11), pp. 7216–7225. doi: 10.1016/j.enpol.2010.07.052.
- Piwowar, H. *et al.* (2018) 'The state of OA: a large-scale analysis of the prevalence and impact of Open Access articles', *PeerJ*, 6, p. e4375. doi: 10.7717/peerj.4375.
- Popp, D. (2016) 'Economic analysis of scientific publications and implications for energy research and development', *Nature Energy*, 1(4), p. 16020. doi: 10.1038/nenergy.2016.20.
- Popp, D., Hascic, I. and Medhi, N. (2011) 'Technology and the diffusion of renewable energy', *Energy Economics*. Elsevier B.V., 33(4), pp. 648–662. doi:

- 10.1016/j.eneco.2010.08.007.
- Suominen, A., Toivanen, H. and Seppänen, M. (2017) 'Firms' knowledge profiles: Mapping patent data with unsupervised learning', *Technological Forecasting and Social Change*. The Authors, 115, pp. 131–142. doi: 10.1016/j.techfore.2016.09.028.
- Tahamtan, I., Safipour Afshar, A. and Ahamdzadeh, K. (2016) 'Factors affecting number of citations: a comprehensive review of the literature', *Scientometrics*, 107(3), pp. 1195–1225. doi: 10.1007/s11192-016-1889-2.
- Tennant, J. P. *et al.* (2016) 'The academic, economic and societal impacts of Open Access: An evidence-based review', *F1000Research*, 5(May), pp. 1–55. doi: 10.12688/f1000research.8460.1.
- Tkaczyk, D. *et al.* (2018) 'Machine Learning vs. Rules and Out-of-the-Box vs. Retrained', in *Proceedings of the 18th ACM/IEEE on Joint Conference on Digital Libraries - JCDL '18*. New York, New York, USA: ACM Press, pp. 99–108. doi: 10.1145/3197026.3197048.
- Viñuales, J. E. (2015) 'The Paris Climate Agreement: an initial examination', *C-EENRG Working Papers*, 9(6), pp. 1–16. doi: 10.2139/ssrn.2704670.
- Wilson, C. *et al.* (2012) 'Marginalization of end-use technologies in energy innovation for climate protection', *Nature Climate Change*. Nature Publishing Group, 2(11), pp. 780–788. doi: 10.1038/nclimate1576.
- Wilson, C. M. (2012) 'Market frictions: A unified model of search costs and switching costs', *European Economic Review*. Elsevier, 56(6), pp. 1070–1086. doi: 10.1016/j.euroecorev.2012.05.007.
- WIPO (2019) *International Patent Classification (IPC)*. Available at: <https://www.wipo.int/classifications/ipc/en/> (Accessed: 9 November 2019).