

The metrics of battery development: Patenting patterns since 2000

Philipp Metzger^a, Sandro Mendonça^b, José A. Silva^c, Bruno Damásio^{a,*}

^a*NOVA Information Management School (NOVA IMS), Universidade Nova de Lisboa, Campus de Campolide, 1070-312 Lisboa, Portugal*

^b*ISCTE Business School, Business Research Unit (BRU-IUL), Avenida das Forças Armadas, 1649-026 Lisboa, Portugal; UECE/REM – ISEG/ University of Lisbon, Rua do Quelhas 6, 1200-781 Lisboa, Portugal; SPRU, University of Sussex, Falmer, Brighton BN1 9RH, UK*

^c*Instituto Dom Luiz, Faculdade de Ciências da Universidade de Lisboa, Campo Grande Edifício C1, Piso 1, 1749-016 Lisboa, Portugal*

Abstract

This study provides a technometric analysis of patent activity on "secondary batteries" from 2000 to 2019. They are rechargeable, redeployable, reusable batteries, hence, inductive of transformative effects in socio-technical systems. The goal is to enrich the existing literature on "green innovation" by providing a deeper insight into the long-run inventive dynamics in energy storage. Mapping and measuring progress is relevant as batteries are a capstone at the intersection of energy and digital transitions, having a key role in the transition to an innovative, closed-loop, circular economy. International patent families are taken as indicator and 92,700 patents were extracted (and text-mined) from the PAT-STAT source. We find that global battery patenting activity accelerated from 2000 to 2019, a sustained development mostly originating in Asia. Four battery types (redox-flow, solid-state, sodium-ion, and lithium-sulfur) have displayed marked innovation in recent years. Furthermore, countries can be clustered in a meaningful way using their patenting performance across these four emerging battery types and the legacy lead-acid technology. Moreover, content analysis on the patents reveals that a number of battery-related technologies and fields of application (such as energy storage systems, battery management systems,

*Corresponding author, bdamasio@novaims.pt

wireless power transmission, electric vehicle charging, and drones) are growing in relevance. Overall, we find evidence of "pro-circularity" in the technologies trajectories developing over time.

Keywords: Secondary batteries, patents, technometrics, text mining

1. Introduction

The importance of batteries has been growing and with two major areas of deployment being electric mobility applications and the storage of energy generated by fluctuating, non-dispatchable, renewable sources. The interlocking role of batteries, a set of technologies at the intersection of the energy-digital transition Yildizbasi (2021), is expected to grow further in the coming decades. A report about the recent developments in electricity storage technologies published by the International Energy Agency (IEA) in association with the European Patent Office (EPO), asserts that under the Sustainable Development Scenario (SDS) defined by the IEA, "the level of deployment and the range of applicability of batteries [...] expands dramatically" (IEA, 2020, p.28). In particular, battery technologies will move beyond consumer appliances and into industrial-size types of equipment: "Charging batteries in electric vehicles will become the largest single source of electricity demand, accounting for around 5% of global demand by 2050" (IEA, 2020, p. 29). Furthermore, "the use of batteries in stationary energy storage applications is [already] growing exponentially" (IEA, 2020, p. 32).

Given the growing awareness regarding battery dynamics, identifying and monitoring the rate and direction of battery innovation as a condition for a low-carbon future is both (analytically) worthwhile and (strategically) urgent. Indeed, battery applications contribute to eliminating pollution from fossil fuels, but they have also systemically transformative effects. Whereas primary batteries are one-off assets, secondary batteries are rechargeable, i.e., these technologies are therefore intrinsically more pro-circular (vis-a-vis primary ones) since they have a longer and more flexible working lifecycle (the energy services ex-

tracted per kilogram of employed material are overwhelmingly superior). Thus, the role of (secondary) batteries in closing loops and building a circular economy is unequivocal (De Jesus et al. (2019), De Jesus et al. (2021)). On the one hand, the recent technological evolution in batteries are enabling new business models and the progressive transition to carbonless energy alternatives, for example, by reducing the carbon burden stemming from fossil-fueled vehicle lock-in and allowing new forms of mobility relying on electric-renewable systems. On the other hand, the development of reliable storage capabilities increases the size of the market for cleaner energies, thus inducing the investment in new infrastructure and the diversification towards natural sources of power. Innovation in batteries themselves, not just their introduction, can also push in the direction of circularity (see De Jesus et al. (2018), Winslow et al. (2018), Zhu et al. (2021), Thompson et al. (2020), Sharma and Manthiram (2020), Li et al. (2022), Wang et al. (2022)). The progressive replacement of single-use batteries for rechargeable ones in consumer electronics has thus been reducing material waste as, in fact, more durable batteries become an important factor as lifetime is key for an increasing variety of use cases. The development of new battery materials and components also allows for tackling another important issue in terms of the circularity, which is reducing/avoiding the use of environmentally hazardous materials such as lead and cadmium, present in the conventional nickel-cadmium and Lead-acid technological generations.

For the present study, we build a new dataset containing 92,700 secondary battery patents (consolidated in terms of *international patent families*, or IPFs) from 2000 to 2019. The raw data was extracted from PATSTAT Online (edition: Autumn 2021), the web interface of the PATSTAT database maintained by the EPO containing a vast collection of data extracted from worldwide patent documents, usable for purposes of statistical analysis (see De Rassenfosse et al. (2014)). Patents emerged in recent years as a crucial raw-material for building an empirically-informed evaluation of progress in tackling pressing global challenges (see Mendonça et al. (2021)), but remain under-exploited in what concerns energy storage developments (Silva et al. (2015); Távora et al. (2020)).

This paper provides a systematic analysis of patent data and patent documentation content for the battery domain with the goal of bringing an empirical indicator forward as a useful tool for operationalising the advance toward transformational societal goals (see Dang and Serajuddin (2020); De Jesus and Mendonça (2018), Morales et al. (2022)).

We find that global battery patenting activity trended upwards in 2000-2019. This stylised fact means that the comparative advantages of secondary approaches (rechargeable, redeployable, reusable batteries) have been continuously on the rise driven by innovation, making a direct contribution to socio-technical circularity. Hence, a better understanding of all aspects of battery evolution is important. We find that the majority of battery patents originates from Asia, and several Asian and European countries exhibit high battery patent intensities in the given time-frame. Four battery technologies (redox-flow, solid-state, sodium-ion, and lithium-sulfur batteries) display increased patenting dynamics from 2000-2009 to 2010-2019. Countries can be clustered in a meaningful way using their patenting distribution over these four emerging battery types. Several battery-related technologies and applications such as energy storage systems, battery management systems, wireless power transmission, electric vehicle charging, and uncrewed aerial vehicles (i.e., drones) grew in relevance both in absolute terms and relative to general battery patenting activity. These results complete and bolster current knowledge regarding the pathways of battery innovation that have been surfacing of late and attracting policy attention (IEA (2020)). The connections of battery innovation with pro-circular transformations may be non-linear, but we find evidence of trajectories of technical change that are less-toxicity intensive, more diverse in the materials employed and more exploratory in the direction of technologies with greater pro-circular potential.

This paper is organized as follows: Section 2 presents the key foundations and concepts relevant to this study. In section 3 the results are presented. These outcomes are discussed in section 4. Section 5 concludes the article. Detailed descriptions of the data selection process and the methods deployed for this analysis are provided in Appendix A.1.

2. Foundations and concepts

2.1. Secondary batteries

90 Secondary batteries are able to receive energy in the form of electricity, store
it, and at a later time—and with a certain loss due to the energy conversion
processes taking place—release it again, feeding electricity back to the grid or
powering a given application. Secondary batteries are rechargeable, unlike pri-
mary batteries that can only discharge once and then need to be discarded. In
95 the context of the ongoing energy transition, away from dispatchable sources
such as coal-fired power plants and towards alternatives such as wind and solar,
whose input is not controllable and hardly synchronous with population and
industry needs, batteries and other means of energy storage constitute a regu-
lating bridge that conjoins the temporal gap between supply and demand while
100 balancing the system as a whole. Moreover, accelerated electrification in the
transporting sector, especially in individual mobility, creates a focusing device
calling out for more batteries with smaller sizes, lighter weight, higher capacity
and longer lifespans (critical technologies have systemic and non-linear impacts
(Marx et al. (2014)). What is more, now in the stationary domain, the emphasis
105 on resilience and energy autonomy has only reinforced the role of batteries as
backup power, in combination of inherently variable sources like solar and wind
(Kosmadakis et al. (2021); Ziegler (2021)).

When referring to batteries, one has to differentiate between the terms “bat-
tery,” “module,” and “cell”. While an entire battery pack potentially consists of
110 multiple modules that are “wired in series and/or (less often) parallel” a module
itself consists of multiple cells that “are connected in series or parallel” (Vezzini,
2014, p. 345). For simplicity’s sake, secondary batteries, meaning battery packs
in their entirety, will hereafter be referred to as *batteries*. The following sub-
section articulates the notion of technical change as a way to encapsulate the
115 phenomenon of knowledge grow and provides a short overview of the advantages
and limitations of using patents as an indicator for measuring it.

2.2. Industrial innovation and the uses of patents as an indicator

Innovation is the process through which ideas and knowledge are converted into useful applications. This means that innovation is a multi-phased process, open to feedback at every stage, molded in an ongoing fashion by a variety of players and institutional settings (Castellaci et al. (2005); Caraça et al. (2009); Ribeiro and Shapira (2020)). As innovation started to be regarded as an empirical phenomenon of significant importance, its measurement became an increasingly topical agenda. Quantification of an intrinsically qualitative process is always a partial approach but it is analytically desirable in order to understand technological change over time and across space; plus, it is valuable in assisting managerial strategy and public policy (Santos et al. (2021)).

An indicator used to capture the notion of innovation is patent data. Typically, interested parties (inventors, owners, intellectual property lawyers, patent offices, etc.) create patents when an invention already is a viable conceptual proposal but not yet operationally tested or fully deployed in practice. Despite only yielding imperfect evidence of innovation, patents are irreplaceable as sources for further study (Mendonça et al. (2019)). When making a case for patents as a proxy for measuring innovation, Zvi Griliches classically explained that patents "are available; they are by definition related to inventiveness, and they are based on what appears to be an objective and only slowly changing standard" (Griliches, 1990, p. 1661). Recently, methodologies incorporating patents and other indicators have been used to relate innovation with environmental progress, human well-being and climate change adaptation (see, e.g., Sovacool et al. (2022)). Indeed, repositioning established empirical tools, such as patents, can be an avenue for such efforts (see, e.g., Losacker (2022), Mendonça et al. (2021)).

Patents are legal documents that describe claims to useful knowledge and assign rights to intellectual property related to inventions. Therefore, they contain information such as geographic locations associated with inventors, descriptions and classifications of the respective inventions, and timestamps like their application and publication dates. This allows for the aggregation of patent counts

alongside geographic, temporal, and technological dimensions and makes them
a suitable material for a myriad of analytical purposes, from competitiveness
studies to sustainability research (Mendonça et al. (2021)) Indeed, patents can
track developments in green innovation, including in strategic emerging sectors
like clean energy (WIPO (2022)). Although they can be seen as a fundamental
direction in a broader pro-sustainability transformation, only a scant literature
can be found drawing on battery patents. The following subsection briefly re-
views it.

2.3. Literature on battery patents analysis

A number of patent-based empirical works have underscored how under-
standing technological potential can inform eco-innovation and climate change
mitigation strategies, including public policies and corporate/start-up develop-
ment efforts. Aaldering and Song (2019) provided an analysis of battery patent
data highlighting developments in post-lithium-ion battery technologies. Mal-
hotra et al. (2021) proposed a citation network analysis combining knowledge
extracted from patent data with results from interviews conducted with lithium-
ion battery experts. Stephan et al. (2019) examined lithium-ion battery patents
from a sectoral diversity perspective and emphasized how the sectoral distance
of prior knowledge affects certain features of subsequent knowledge. Kittner
et al. (2017) and Ziegler and Trancik (2021) employed the patent proxy in their
efforts to model the forces driving the prices of lithium-ion batteries, and found
that cumulative patent filings is the best predictor of real prices scaled by energy
capacity.

The current analysis extends the existing literature on this matter. Specifi-
cally, it aims to confirm and consolidate the findings presented in the IEA and
EPO report (IEA (2020)). It can be understood as a continuation of their basic
methodological approach, enriched by some reasonable additions, which allow
for a more granular perspective on some aspects of this topic.

The IEA and EPO report presents information extracted from patents re-
lated to batteries and electricity storage. Our own study is narrower (focuses

on battery technology only) but has a longer time span. The research gaps that we identified and which the current study aims to fill are how patent counts
180 are distributed across continents, how scaling them by the sizes of the respective labor forces affects the outcome of the analysis, what their distribution across another technological classification scheme looks like, how countries can be characterized based on their position in a resulting technology space, and what information can be extracted from patent titles and abstracts.

185 The authors of the report by IEA and EPO use the concept of international patent families (IPF) for aggregating and counting patent applications. They claim that an IPF "is a reliable proxy for inventive activity because it provides a degree of control for patent quality by only representing inventions for which the inventor considers the value sufficient to seek protection internationally" ((IEA, 2020, p.4)).
190 The following subsection explains the differences between patent applications, patent families, and international patent families and highlights some advantages and limitations of the latter concept.

2.4. *International patent families*

A patent application is a formal request made by one or several applicants
195 at a patent office of their choice. These could be the European Patent Office (EPO), the United States Patent and Trademark Office (USPTO), or any other national or regional patent office. The applicants' goal is to obtain legal protection for an invention that they deem (1) directed to patentable subject matter, (2) novel, (3) inventive, and (4) capable of industrial application, which
200 are the four conditions for patentability (OECD (2009)). The term *patent family* refers to the whole set of patent applications covering the same invention (Dechezleprêtre et al. (2017)). By counting patent families instead of individual applications, double-counting of inventions is avoided. By restricting the search protocol the scope to only patent families that contain applications filed in two
205 or more countries, one obtains international patent families. The benefit of this restriction is that only patents of higher expected value are assessed, resulting in a more homogeneous dataset with better comparability between elements.

As Schmoch and Gehrke (2022) discussed, three limitations regarding the concept of IPFs should be considered: First, the propensity to patent in foreign territories differs between countries of origin, meaning that, for example, an applicant from a European country might be more inclined to seek protection in another European country than an applicant from China might be inclined to seek protection in the US. This can be problematic because both situations would imply that the respective patent is filed in two countries, thus making their patent family an international patent family. Secondly, in specific technologies, the patent numbers for some countries, such as Japan, may be overestimated. Thirdly, there can be some turbulence in the evidence since IPFs with seemingly two members at the stage of applications can be reduced to one member, later on, something that may happen with Chinese inventors (regarding the Chinese case, we further refer to Frietsch and Kroll (2020)). Schmoch and Gehrke (2022) discuss several other concepts that exist parallel to IPFs, highlighting their advantages and limitations. For comparability to the IEA-EPO report, we have kept international patent families as our frame; therefore, all depicted counts refer to IPFs.

2.5. International Patent Classification (IPC)

The international patent classification system (IPC) provides a hierarchical classification scheme that categorizes patents according to different technological areas. This study builds on patents that can roughly be characterized in the following way: (1) innovations related to the casing, wrapping, or covering, i.e., non-active parts of batteries; (2) innovations in battery electrode manufacturing; (3) innovations related to the manufacturing process of secondary cells; and (4) innovations related to charging of batteries. Patents belonging to these four fields were identified using the IPC classification scheme, which is a constituent part of the data provided at the PATSTAT database. For more details regarding data acquisition and processing see Appendix A.1.

3. Results

3.1. Basic stylized facts

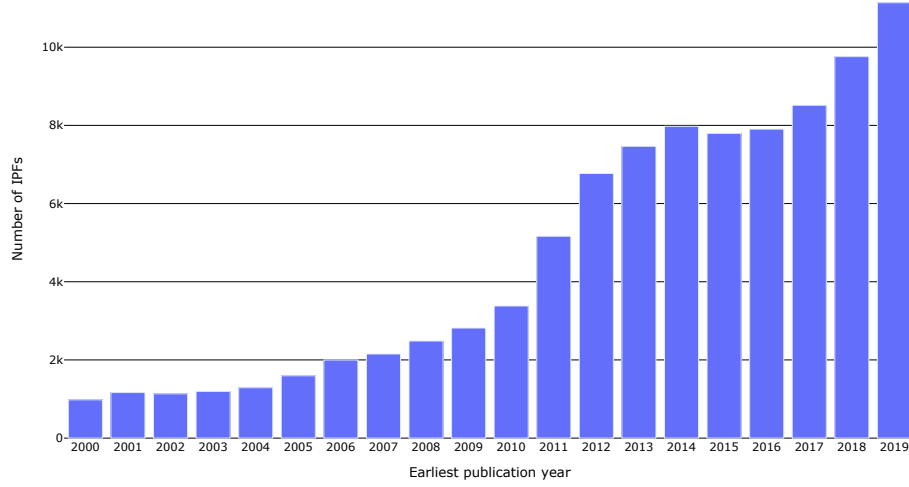


Figure 1: Total number of battery IPFs, 2000-2019.

This section presents the results obtained by counting battery IPFs both on global and national levels, disaggregating the patterns into different technologies, and analyzing how countries position themselves in the technology space. Furthermore, we report the results of a text mining approach deployed on patent abstracts. This first subsection contains some basic patterns highlighting the most essential battery patenting trends.

The global aggregate yearly volume of battery IPFs increased almost every year during the timeframe assessed in this study. There were slight decreases only for two pairs of adjacent years: from 2001 to 2002 and from 2014 to 2015. The whole time period’s average yearly growth rate in battery IPFs is 14.30% (all percentages are rounded to two decimal places) so that between 2000 and 2019 the total IPF output increased more than 11-fold. This dynamic is displayed in Fig. 1.

Asian countries dominate the battery market: The Asian continent’s mean annual battery IPF output is approximately four times higher than Europe and

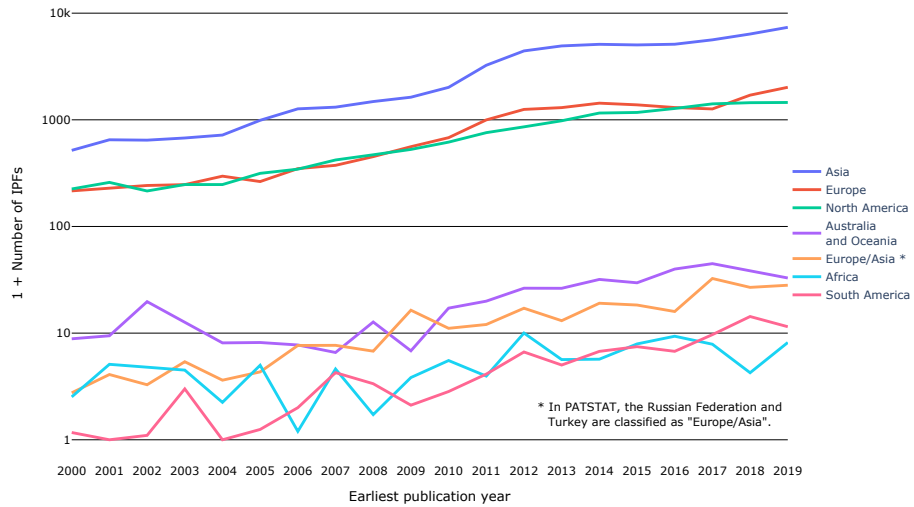


Figure 2: Battery IPFs by inventors' continents of origin, 2000-2019.

Note: The y-axis is log-scaled and all values are incremented by 1. It is clear that the number of battery IPFs from Asia (blue) is considerably higher than that of any other continent.

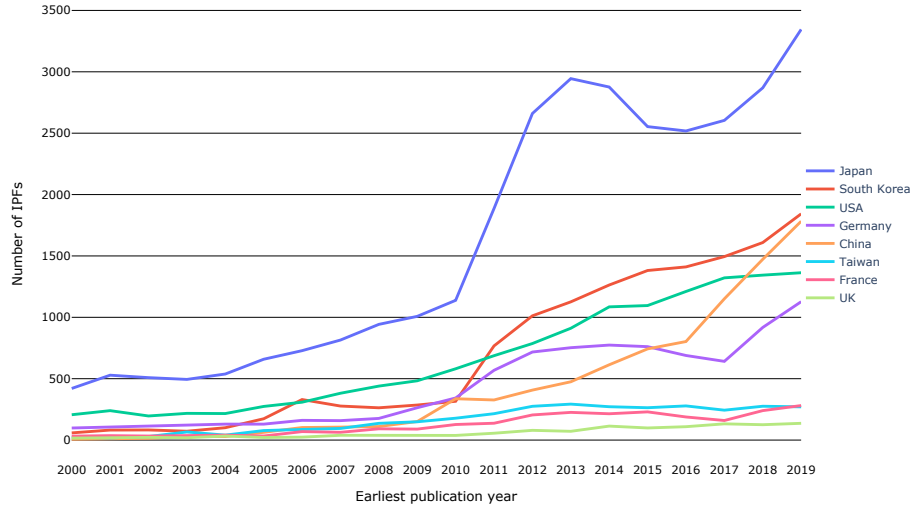


Figure 3: Battery IPFs by inventors' countries of origin, 2000-2019.

Note: The eight countries with the highest total battery IPF counts over the given timeframe are displayed. Japan (blue) has the highest battery IPF output in the given timeframe, whilst other countries' IPF counts (especially South Korea's (red) and China's (orange)) have been surging in the recent decade.

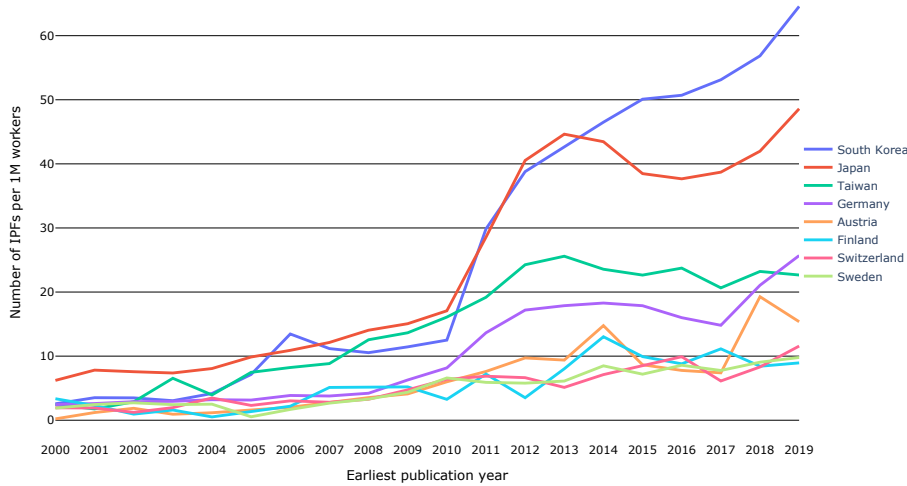


Figure 4: Battery IPFs per 1M workers by inventors' countries of origin, 2000-2019.

Note: The eight countries with the highest total battery IPF intensities over the given time-frame are displayed. In this perspective, South Korea (blue) overtook Japan (red) in 2014.

North America's (factor 3.57 and 4.10, respectively). Furthermore, the number of IPFs from Asia increased by 15.96% on average each year during the 2000-2019 period. The average increase values for Europe and North America are 13.46% and 10.80%, respectively (see Fig. 2; log-scaled y-axis).

Breaking down battery IPF counts by inventors' countries of origin, the dominance of Asia becomes even more apparent. Figure 3 shows the eight countries with the highest total battery IPF output over the whole timespan. In 2019 the three top countries in terms of battery IPF output were from the far east: Japan, South Korea, and China. These were followed by the US, Germany, France, Taiwan, and the UK. Japan, the undisputed leader in battery IPF counts during the whole time frame, has been displaying a vibrant rate in the dynamics of inventive output since 2016. China is catching up fast with South Korea, which has held second place in battery IPF output since 2011 when it surpassed the US (for the Chinese case see Hsu et al. (2021)). Germany also displays an upward trend in battery IPF output. Please note the similarity of trajectories presented in this plot to those depicted in Figs. 6.2 and 6.3 of the

IEA and EPO report (IEA (2020)). The higher numbers in the current study
 270 result from the underlying data being defined somewhat differently. The most
 significant difference is that IPFs related to the field of battery charging were
 included in the current study’s dataset.

By scaling the numbers shown in the previous plot by each country and
 year’s labor force count, one obtains battery IPF intensities (Neuhäusler et al.
 275 (2019)). This measure gives the viewer a different perspective on the IPF counts,
 allowing for the assessment of a country’s innovative output relative to the size
 of its working population. Figure 4 shows the eight countries with the highest
 scaled total battery IPF output over the whole period and it can be seen that
 in contrast to Fig. 3, some small European countries are stepping up: Austria,
 280 Finland, Switzerland, and Sweden are part of the top eight. It is also worth
 noting that, in this light, South Korea overtook Japan in 2014, establishing
 itself as the global leader in terms of battery patent intensities.

3.2. Battery technologies

By assigning battery technology sub-areas to patent families a disaggrega-
 285 tion of the dataset into 19 battery cell technologies was obtained. This pro-
 cess is described in detail in subsection A.1.2. The technology classes used in
 this study are "lead-acid," "lithium-air," "lithium-ion," "lithium-sulfur," "solid-
 state," "other lithium," "magnesium-ion," "nickel-cadmium," "nickel-iron," "nickel-
 zinc," "nickel-metal hydride," "Rechargeable alkaline," "sodium-sulfur," "sodium-
 290 ion," "aluminum-ion," "calcium(-ion)," "organic radical," "redox-flow," and "nickel-hydrogen."

Figure 5 presents the developments of IPF counts in the eight major cat-
 egories. They were selected based on their total IPF count in the entire time
 frame of 2000-2019. While the number of IPFs related to lead-acid batteries
 (i.e. arguably the least circular of the technological options) has been rel-
 295 atively stable over the depicted 20 years, which resulted in its overall share
 in battery IPFs decreasing steadily over this time period, and while recharge-
 able alkaline batteries exhibit a slight downwards trend, lithium-ion batteries
 and other lithium-based battery technologies have soared drastically. Less rele-

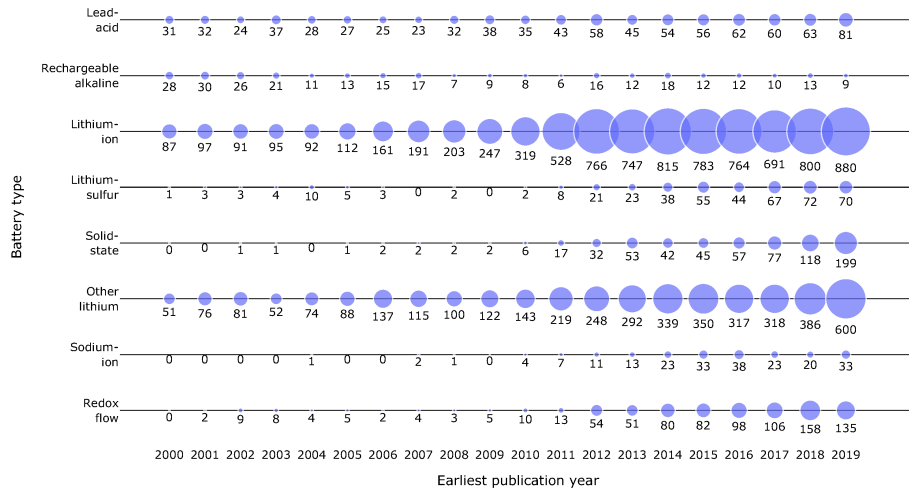


Figure 5: Global battery patenting activity for the major battery types, 2000-2019.

The depicted battery IPF fractional counts are rounded to the closest integer. The eight technologies with the highest total battery IPF count over the given timeframe are displayed. The relative importance of lead-acid and rechargeable alkaline batteries decreased, whilst IPF counts for lithium-ion batteries and other lithium-based battery technologies have soared robustly. After 2010 four technologies emerge: Lithium-sulphur, solid-state, sodium-ion, and redox-flow batteries.

vant today than lithium-ion batteries but with considerably higher counts than
 300 other smaller battery technologies are the four remaining categories presented in
 Fig. 5: Patenting activity related to lithium-sulfur, solid-state, sodium-ion, and
 redox-flow batteries have seen a notable increase in IPF counts in 2010-2019.
 In 2019 solid-state batteries reached an all-time maximum.

It must be mentioned that solid-state batteries are specific a specific config-
 305 uration of batteries, that is, with solid electrolytes and electrodes, mostly im-
 plemented in the framework of lithium-ion solutions. In that sense, one might
 assert that the emergent redox-flow, lithium-sulphur and sodium-ion technolo-
 gies provide a more substantial contribution to technological heterogeneity and
 can lead to higher diversification of the materials used in battery manufacture
 310 thus avoiding the overexploitation of scarce resources available in nature such
 as the those already extensively used in the dominant lithium-ion technologies
 (like lithium and cobalt). In this sense, the increase in technology diversity
 promoted by innovation has the potential to increase the overall circularity of
 battery development.

315 The observation that the recent decade displayed increased patenting activity
 in these four emerging technologies motivates the way the next part of the
 analysis is set up: The following subsection describes the results obtained by
 clustering countries based on their position in a technology space computed
 using their technology distribution of the years of 2010-2019 (6).

320 3.3. Clustering

The most suitable technology space for clustering was found to be spanned
 by the countries' distribution values over the four emerging technologies lithium-
 sulfur, solid-state, sodium-ion, and redox-flow, which display increased patent-
 ing activity after 2010, alongside the older lead-acid technology.

325 In attempting to cluster 36 countries using data from 2010 to 2019 k-means
 was found to be the algorithm with a better R^2 value for all relevant numbers
 of clusters (for details on this metric see Appendix A.1). Setting the numbers of
 clusters to two, we obtained a clear separation of the dataset between countries

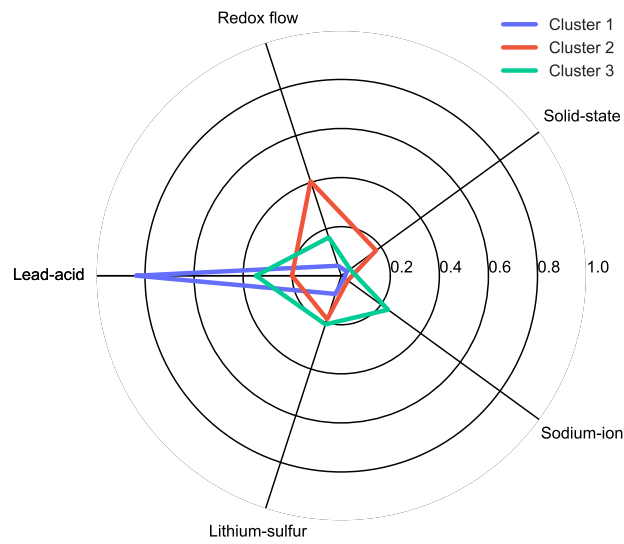


Figure 6: Cluster of inventors' countries of origin, 2010-2109.

Note: While countries from cluster 1 are more focused on lead-acid batteries, clusters 2 and 3 exhibit a higher patenting activity related to the four emerging technologies of redox-flow and solid-state batteries (cluster 2) and lithium-sulphur and sodium-ion batteries (cluster 3). K-means algorithm.

with a high focus on lead-acid batteries (81.91% of IPFs are related to lead-acid
batteries in this cluster) and countries with comparatively high shares of IPFs
related to the four emerging technologies and consequently a relatively low share
of lead-acid related IPFs (19.55%).

Setting the number of clusters to three to achieve a more granular separation,
one finds that the lead-acid focused cluster from the previous stage is still fairly
intact, while the "emerging technologies" cluster has been separated into two.
This division results in one cluster displaying a stronger focus on redox-flow and
solid-state batteries and another exhibiting a higher relative focus on sodium-
ion and lithium-sulfur-related IPFs. Figure 6 shows the distribution profiles of
the three-clusters solution generated with the k-means variable "random_state"
set to zero. The variable "random_state" determines the centroid initialization
of k-means and results in deterministic runs of the algorithm when a value is
assigned to it.

While the approximate shape of the clustering profile depicted in Fig. 6 is
fairly insensitive to alterations or non-assignment of "random_state," the affilia-
tion of the countries to their clusters varied enough to motivate running k-means
a higher number of times (with the variable "random_state" undefined) to com-
pute each country's cluster affiliation distribution for assessing which cluster
each country belongs to in the majority of events. Running k-means 10,000
times resulted in the following most probable cluster affiliations:

- Cluster 1 (16 countries):

India , Turkey , Russia , Bulgaria , New Zealand , Luxembourg , Poland ,
Sweden , Mexico , Malta , North Korea , Serbia , Greece , Hungary ,
Kazakhstan , Israel .

- Cluster 2 (13 countries):

USA , Germany , Taiwan , Austria , Netherlands , Thailand , Switzerland ,
South Korea , Japan , Belgium , Italy , Australia , Hong Kong .

- Cluster 3 (7 countries):

Canada, Spain, Ukraine, UK, France, Norway, China.

Inside each cluster, countries are ordered by (1) their probability p to be in this cluster, and (2) their total IPF count in the five categories. Each country's name is colored according to the following schema, indicating its probability p to belong to the respective cluster:

$p = 1$ $p \in [0.99, 1)$ $p \in [0.9, 0.99)$ $p \in [0.5, 0.9)$

A value of $p = 1$ indicates that a country was assigned to this cluster during each of the 10,000 runs, meaning that its cluster affiliation appears to be quite insensitive to the algorithm's centroid initialization.

In terms of circularity, in line with what was mentioned in the previous section, we can assert that due to their higher technological diversity Clusters 2 and 3 have the potential to provide a higher contribution to a more Circular Economy than Cluster 1, which is mainly focused on Lead acid technologies. In comparing Cluster 2 and 3, it stands out that Cluster 2, while having a strong emphasis on solid-state batteries (which as mentioned is essentially a particular type of lithium-ion battery), is mainly focused on two emerging technologies (redox-flow and lithium-Sulphur). In contrast, Cluster 3 reveals robust innovation activity in three emerging technologies outside the lithium-ion technologies framework (i.e., redox-flow, lithium-sulphur and sodium-ion), suggesting that Cluster 3 has could have a higher potential to contribute to circularity in the future.

3.4. Patent title and abstract mining

The content material of patents is relevant evidence that can be mined, processed, and sorted to leverage classic patent analysis (Hsu et al. (2020)). Two methods were implemented in order to use patent wordage as empirical material. Both patent abstracts and titles were searched for meaningful phrases. The first approach, henceforth called *n-gram counts* (Figure 7), was to simply count occurrences of n-grams in patent abstracts and titles for each year. The second approach, from now on referred to as *n-gram intensities* (Figure 8), was to additionally scale these counts by the respective year's number of abstracts or titles,

	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
electrode active material	89	82	72	121	136	157	241	278	294	356	427	991	1253	1288	1323	1386	1273	1308	1382	1990
active material layer	17	9	29	27	37	96	114	143	134	188	186	364	419	553	499	383	336	510	619	883
lithium secondary battery	46	91	72	65	112	128	192	140	124	178	222	340	451	487	490	410	492	374	540	844
lithium ion battery	15	26	41	31	16	49	58	83	99	121	197	321	476	421	449	520	618	568	702	685
energy storage device	5	41	27	39	33	20	44	70	109	94	147	183	224	367	367	277	370	538	576	648
secondary battery electrode	29	27	21	25	55	82	113	109	95	122	146	286	417	422	384	400	419	441	435	593
electrode current collector	7	7	10	12	15	15	24	46	57	104	96	135	206	171	155	172	215	226	412	539
power storage device	11	3	7	10	2	8	4	30	69	37	183	166	317	310	361	311	311	520	293	394
electrolyte secondary battery	25	27	41	36	65	74	110	118	180	129	113	214	301	269	437	425	412	286	264	405
non-aqueous electrolyte secondary	24	38	45	46	67	75	109	116	182	135	121	216	307	288	447	425	419	284	248	384
plurality battery cell	3	1	3	1	7	9	10	28	26	29	32	135	175	229	177	206	212	232	314	351
ion secondary battery	10	25	15	23	32	58	68	73	75	81	162	261	331	409	428	468	388	337	300	357
power supply device	11	3	25	7	20	28	35	28	84	60	80	140	227	259	285	222	228	256	281	348
lithium ion secondary	12	29	18	30	33	63	72	74	76	84	164	270	366	418	442	506	379	342	291	345
current collector electrode	5	8	6	4	8	11	12	27	37	45	56	127	113	114	138	97	137	138	217	323
active material lithium	17	24	16	28	27	36	36	43	48	65	100	168	256	183	225	255	222	177	160	266
cathode active material	6	10	23	31	9	28	59	49	44	64	92	79	146	168	199	182	168	198	211	247
power supply system	7	20	25	18	27	21	32	52	47	59	82	150	187	162	165	145	166	185	218	241
energy storage system	0	7	0	8	2	11	6	25	35	32	46	93	80	116	121	106	138	165	200	222
electrode mixture layer	0	0	0	3	0	0	4	14	25	38	26	56	87	93	89	111	151	98	170	219
solid electrolyte layer	5	2	1	0	3	0	3	3	19	11	21	48	46	92	43	70	68	105	126	208
solid state battery	0	0	0	1	0	0	1	2	5	4	3	11	29	58	49	29	33	53	95	198
layer electrode active	3	1	2	3	3	5	6	17	12	20	30	60	71	93	85	112	102	105	120	200
battery management system	0	3	3	1	2	3	11	30	39	20	22	61	91	113	97	147	107	139	163	185
material layer electrode	2	0	1	3	3	11	15	19	9	30	34	76	79	89	82	86	74	101	91	178
energy storage unit	1	1	21	1	4	7	20	16	26	47	43	81	78	78	44	94	92	101	201	176
secondary battery lithium	6	15	4	10	8	13	10	13	21	25	40	46	69	69	90	96	106	70	125	177
anode active material	6	2	4	15	14	20	58	61	65	55	77	80	92	113	211	108	74	117	166	175
transition metal oxide	3	7	3	9	3	10	20	18	13	28	38	49	54	68	91	132	106	80	83	167
active material particle	4	6	12	6	14	19	41	32	32	36	38	57	94	93	93	150	161	213	199	164
collector electrode active	1	2	4	1	3	6	4	15	13	15	28	68	44	92	109	67	67	72	82	157
active material electrode	27	23	13	15	20	20	32	34	37	40	52	107	119	126	118	161	128	131	123	182
electrical energy storage	1	3	10	15	18	3	16	38	15	28	52	32	89	116	74	77	79	100	125	153
wireless power transmission	0	0	0	0	0	0	0	0	2	13	23	33	115	105	183	172	154	223	143	148
redox flow battery	0	0	8	12	2	1	0	0	0	2	2	22	44	30	50	93	86	102	120	146
power storage element	1	0	0	0	0	0	0	0	13	11	10	4	46	58	93	112	160	162	170	143
material electrode active	7	4	4	5	5	14	11	18	17	33	27	50	82	78	69	84	89	70	84	146
material lithium ion	5	10	3	13	5	17	12	18	23	33	63	106	162	134	153	181	142	125	126	140
power transmission device	0	0	0	2	0	0	0	0	2	30	27	41	53	92	121	173	96	113	136	134
lithium transition metal	8	16	6	19	24	31	20	19	36	29	42	74	75	82	106	138	112	87	112	138
battery module plurality	1	4	0	2	0	5	17	3	15	15	18	52	62	71	65	71	60	64	117	127
non-aqueous electrolyte solution	3	9	5	0	11	13	12	12	22	14	31	42	62	61	70	45	54	37	68	125
material lithium secondary	11	22	21	21	33	32	27	27	26	50	54	81	125	105	134	99	102	80	84	130
electric vehicle charging	0	0	0	0	0	0	1	0	5	9	25	48	93	113	65	53	38	50	64	118
battery electrode active	5	7	9	13	21	14	22	44	25	36	39	77	128	120	115	133	111	133	102	122
power supply circuit	10	12	27	14	11	25	12	44	38	26	31	59	71	59	73	37	63	101	100	127
battery cell electrode	0	1	4	1	1	2	2	8	2	11	14	24	30	59	42	43	53	47	72	117
control unit configured	0	0	0	0	0	2	1	2	6	6	9	13	34	34	42	46	50	62	61	114
state secondary battery	0	4	2	1	3	3	4	5	9	9	16	23	31	41	28	27	57	96	45	112
electrode lithium secondary	4	8	11	10	17	19	8	13	17	19	29	39	31	48	35	28	38	39	74	114

Figure 7: Trigram occurrences in battery patent abstracts.

Note: The top 50 trigrams in terms of their count increase between 2000 and 2019 are displayed. The color gradients represent intra-row relationships.

	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
electrode active material	91	70	64	102	106	100	123	129	119	127	126	192	185	173	167	179	163	155	143	181
active material layer	17	8	26	23	29	61	58	67	54	67	55	71	62	74	63	50	43	60	64	80
energy storage device	5	35	24	33	26	13	23	33	44	33	44	36	33	49	46	36	47	64	60	59
lithium ion battery	15	22	36	26	12	31	30	39	40	43	58	62	70	56	57	67	79	67	73	62
electrode current collector	7	6	9	10	12	10	12	21	23	37	28	26	30	23	20	22	27	27	43	49
lithium secondary battery	47	78	64	55	87	82	98	65	50	63	66	66	67	65	62	53	63	44	56	77
plurality battery cell	3	1	3	1	5	6	5	13	10	10	9	26	26	31	22	27	27	27	32	32
power storage device	11	3	6	8	2	5	2	14	28	13	54	32	47	42	45	40	40	62	30	36
current collector electrode	5	7	5	3	6	7	6	13	15	16	17	25	17	15	17	13	17	16	22	29
secondary battery electrode	30	23	19	21	43	52	58	51	38	43	43	56	62	57	48	52	54	52	45	54
ion secondary battery	10	21	13	19	25	37	35	34	30	29	48	51	49	55	54	61	50	40	31	32
power supply device	11	3	22	6	16	18	18	13	34	21	24	27	34	35	36	29	29	30	29	32
energy storage system	0	6	0	7	2	7	3	12	14	11	14	18	12	16	15	14	18	20	21	20
electrode mixture layer	0	0	0	3	0	0	2	7	10	14	8	11	13	12	11	14	19	12	18	20
lithium ion secondary	12	25	16	25	26	40	37	34	31	30	49	52	54	56	56	65	48	41	30	31
solid state battery	0	0	0	1	0	0	1	2	1	1	2	4	8	6	4	4	6	10	18	18
battery management system	0	3	3	1	2	2	6	14	16	7	7	12	13	15	12	19	14	16	17	17
cathode active material	6	9	20	26	7	18	30	23	18	23	27	15	22	23	25	24	21	23	22	22
layer electrode active	3	1	2	3	2	3	3	8	5	7	9	12	11	12	11	14	13	12	12	18
energy storage unit	1	1	19	1	3	4	10	7	10	17	13	16	12	10	6	12	12	12	21	16
power supply system	7	17	22	15	21	13	16	24	19	21	24	29	28	22	21	19	21	22	23	22
material layer electrode	2	0	1	3	2	7	8	9	4	11	10	15	12	12	10	11	9	12	9	16
solid electrolyte layer	5	2	1	0	2	0	2	1	8	4	6	9	7	12	5	9	9	12	13	19
wireless power transmission	0	0	0	0	0	0	0	0	1	5	7	6	17	14	23	22	20	26	15	13
redox flow battery	0	0	7	10	2	1	0	0	0	1	1	4	7	4	6	12	11	12	12	13
collector electrode active	1	2	4	1	2	4	2	7	5	5	8	13	7	12	14	9	9	9	8	14
electrical energy storage	1	3	9	13	14	2	8	18	6	10	15	6	13	16	9	10	10	12	13	14
power transmission device	0	0	0	2	0	0	0	0	1	11	8	8	8	12	15	22	12	13	14	12
transition metal oxide	3	6	3	8	2	6	10	8	5	10	11	10	8	9	11	17	14	9	9	15
power storage element	1	0	0	0	0	0	0	0	5	4	3	1	7	8	12	14	20	19	18	13
electrolyte secondary battery	26	23	36	30	50	47	56	55	73	46	33	42	45	36	55	55	53	34	27	37
active material particle	4	5	11	5	11	12	21	15	13	13	11	11	14	12	12	19	21	25	21	15
electric vehicle charging	0	0	0	0	0	0	1	0	2	3	7	9	14	15	8	7	5	6	7	11
battery cell electrode	0	1	4	1	1	1	1	4	1	4	4	5	4	8	5	6	7	6	7	11
battery module plurality	1	3	0	2	0	3	9	1	6	5	5	10	9	10	8	9	8	8	12	12
control unit configured	0	0	0	0	0	1	1	1	2	2	3	3	5	5	5	6	6	7	6	10
non-aqueous electrolyte secondary	25	33	40	39	52	48	56	54	73	48	36	42	45	39	56	55	54	34	26	35
state secondary battery	0	3	2	1	2	2	2	2	4	3	5	4	5	5	4	3	7	11	5	10
secondary battery lithium	6	13	4	8	6	8	5	6	8	9	12	9	10	9	11	12	14	8	13	16
anode active material	6	2	4	13	11	13	30	28	26	20	23	16	14	15	27	14	9	14	17	16
current collector layer	0	0	1	2	2	1	2	0	0	0	1	1	1	5	2	1	2	4	10	9
unmanned aerial vehicle	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	4	13	16
electrode active substance	1	2	9	0	14	10	3	1	1	2	4	4	7	6	11	7	8	4	7	9
plurality battery module	1	4	1	6	1	2	4	6	1	4	3	10	7	7	6	4	8	5	11	9
non-aqueous electrolyte solution	3	8	4	0	9	8	6	6	9	5	9	8	9	8	9	6	7	4	7	11
power receiving device	0	0	0	0	2	3	0	3	1	12	6	8	11	16	22	14	10	10	12	8
present electrode active	0	2	2	2	1	1	1	2	1	2	2	7	5	3	7	7	9	8	7	8
solid state secondary	0	0	0	0	0	0	1	1	0	0	1	2	1	2	1	1	6	8	3	8
solid electrolyte material	0	0	0	0	0	0	0	0	2	0	2	3	4	5	3	1	2	4	4	8
power storage system	0	0	1	0	0	1	0	4	0	1	1	4	7	7	10	6	9	4	8	8

Figure 8: Trigram occurrence intensities in battery patent abstracts.

Note: The unit of measure is occurrences per 1,000 abstracts and all values are rounded to the closest integer. The top 50 trigrams in terms of their intensity increase between 2000 and 2019 are displayed. The color gradients represent intra-row relationships.

respectively. In this case, the resulting unit of measure for n-gram intensities is occurrences per 1,000 abstracts or titles, respectively, and all depicted n-gram intensities are rounded to the closest integer. More precisely, each cell displayed in Figure 8 is the respective occurrence count (from Figure 7) divided by a scale factor (the number of abstracts associated with the respective year $\times 1,000$) and then rounded to the closest integer. The reason is that some patent families do not have a non-NaN English abstract. That is, the number of abstracts associated with a year can be lower than the number of patent families associated with that year. For purposes of sensitivity analysis, unigrams (single words), bigrams (strings containing two words), and trigrams (arrays with three words) were extracted and processed. The resulting n-gram counts and n-gram intensities were sorted in three different ways, which are described in detail in subsection A.1.6 (Appendix A.1). The results that we found most meaningful and thus selected for presentation in this paper are the top 50 increasing trigrams extracted from battery patent abstracts. The terms are displayed in descending order of total increase over the given 20-years time period in Figures 7 (trigram counts) and 8 (trigram intensities). A holistic analysis is provided, considering the different analyses that were implemented (but not shown here for the sake of space).

Trigram counts display several expectable trends like the surge of "lithium secondary battery" and "lithium ion battery." The occurrence counts for these two trigrams increased from 46 to 844 and from 15 to 685, respectively, between 2000 and 2019 and the trigram intensities of "lithium ion battery" indicate a robust upward dynamic not only in absolute terms but also relative to battery patenting activity. The increase of the term "energy storage system," which is also confirmed by its intensity's trajectory, hints at an upsurge in the importance of increasingly complex systems for managing energy storage. This is buttressed by the term "battery management system," also occurring in both counts' and intensities' top 50 trigrams. As already established by Fig. 5, solid-state batteries have been growing in relevance, especially in the past decade. The increasing counts and intensities for the terms "solid electrolyte layer" and "solid state battery" after 2010 confirm this. Notable trigrams in the sub-

fields of battery charging and electric vehicles are "wireless power transmission" and "electric vehicle charging," which have both increased considerably in
420 both counts and intensities. The surge in relevance for redox-flow batteries (see Fig. 5) is also confirmed by both counts and intensities ("redox-flow battery"). The trigrams "plurality battery cell" (results from "plurality of battery cells" due to stop word removal and lemmatization) and "battery module plurality"
425 (both present in counts and intensities) hint at a substantial increase in innovative output related to compositions of cells and modules inside battery packs. An unexpected, yet reasonable, appearance in the top 50 trigram intensities is the term "unmanned aerial vehicle," exhibiting 4, 13, 16, and 8 occurrences per 1,000 abstracts in 2016, 2017, 2018, and 2019, respectively. It indicates an
430 increased field of application related to the deployment of battery technology in drones.

The relevance of innovation in batteries for appraising the transition to a circular economy can be further discussed by analysing these text materials. Indeed, as gleaned from Figure 7, 15 of 50 trigrams with a more significant growth
435 in the period 2000-2019 have references to secondary, rechargeable or storage. That is, the technological paradigm is not about primary (less reusable, less enduring, less re-deployable approaches). Moreover, these same (pro-circularity) descriptors appear on average 4.15 times in the top 10 trigrams over the period thus providing suggestive evidence on the pro-circularity of battery innovation
440 trends.

Another sign of transformative innovation emerges from this content analysis. The frequent appearance of references to "hydrogen absorbing" "alloy absorbing", "nickel hydroxide", and "hydrogen storage" in the beginning of the time series (mostly in the years 2000, 2001, and 2002) might be attributed to
445 the innovation effort to find alternatives to nickel-cadmium batteries types by replacing the highly toxic cadmium by substitutes based on nickel-metal hydride. In other words, in the early part of the first decade there is evidence on breaking new ground towards cleaner combinations, less toxic materials, more earth/ocean-friendly solutions.

450 The trigram analysis overall confirms the results obtained in section 3.2,
 in particular the prominence of the Lithium-ion technologies and the nature
 of most relevant alternative technological paths. But it also hints at the non-
 linearity of progress toward safer and more sustainable forms of energy storage.
 Two undercurrents of technical change are particularly telling in this respect,
 455 namely the rising importance of non-aqueous electrolytes and the growing inter-
 est in solid state batteries (both mainly associated to the lithium-ion batteries).
 These trends have a rather complex relation with the circular economy. Non-
 aqueous (solid) electrolytes tend to be made of more toxic materials than aque-
 ous ones. Plus, solid state batteries, that is, batteries with solid electrolytes,
 460 tend to have shorter lifecycles. So, at first glance both trends are going against
 circularity principles. However, both approaches allow for the increase of the
 energy density of batteries, a feature that is crucial to improve the performance
 of electric cars, making them more appealing to users, thus accelerating the
 transition away from the fossil fuel powered cars to the electric ones, thus im-
 465 proving circularity at a systemic level. In other words, it may well be that some
 micro-heuristics (going for non-aqueous electrolytes and solid state batteries),
 which in themselves are less circular, can have pro-circular effects at a macro-
 systemic level of analysis. Hence, technology analysis and patent indicators are
 only a partial and subsidiary approximation to the broader meaning of battery
 470 innovation and its links with the evolving socio-technic system.

It should be noted that the results obtained by processing the data through
 all method combinations are not presented in this paper (i.e., application titles
 and other sorting methods), they can best be appraised by accessing the file
 "03.title_and_abstract_mining.html," which is part of the GitHub repository as-
 475 sociated with this study. The said repository can be accessed by following the
 link provided in the Appendix A.1.

4. Discussion

Examining Fig. 1, one could infer that the inflection point between 2011 and 2012 may result from the global financial crisis and the subsequent recession. Assessing Fig. 1, Fig. 3, and Fig. 4 jointly, one can identify a clear difference in annual battery patenting activity between the two decades assessed in this study (2000-2009 and 2010-2019), both on a global level and for several countries. Combining this knowledge with Fig. 2, it becomes clear that Asia drives the major part of the increase in battery patenting activity.

The observation obtained from Fig. 2 that the Asian continent has by far the highest battery IPF output worldwide should be accompanied by the remark that the countries classified as "Asia" in PATSTAT account for approximately 60% of the world's labor force. Moreover, when computing each continents' battery IPF intensities, one observes that Asia falls behind both Europe and North America. For interested readers, IPF intensities for each continent are displayed in Fig. B.10 in the Appendix A.2.

Concerning the country-wise patent counts and intensities presented in Fig. 3 and Fig. 4, it is worthwhile mentioning that comprehensive analyses undertaken before defining the final dataset resulted in the observation that most battery patent applications from China in the considered time frame of 2000-2019 are only filed nationally. Given the IPF constraint deployed for this study and the IEA and EPO report (IEA (2020)), these solely nationally filed applications are not considered in either one. In fact, in the current study's dataset, IPFs make up only 19.41% of all battery patent families. It is reasonable to define the data for the current study as such (the same for the recent analysis undertaken by IEA and EPO) because it can be expected that patents filed in only one country are of considerably lesser "value" than international patent families. Including them would thus result in a rather inhomogeneous dataset. Nonetheless, it is worth noting that if the IPF restriction was to be discarded and one-country patent families were to be considered, China (which in fact is the world's largest producer and market) would take the first place in battery patent counts in

the majority of years of the recent decade. As a resulting thought, it would be worthwhile investigating the battery patenting dynamics of China in detail within the context of future research to shed light on why China’s battery patenting behavior is so nationally-focused and what implications this has for technology analyses in this field.

This study found robust country clusters as they advance along emergent battery innovation pathways. This outcome means there is country variation in terms of technological capabilities and strategies. We are thus witnessing specialization and heterogenous technological trajectories regarding this dimension of the energy transition. These different profiles and choices may be non-neutral in terms of circularity potential. By interpreting the clustering solution presented in subsection 3.3, the three resulting clusters could be characterized as follows:

- Cluster 1 – Lead-acid based:

Many of these countries’ battery innovation results are made up of lead-acid battery patents. Their share of battery patents related to the four analyzed emerging technologies are close to zero, except for their lithium-sulfur component, which accounts for approximately 8% of their IPF output in 2010-2019. This ”insurgent” cluster contains countries like India, Turkey, and Russia that are considerably industrialized but are not known for their innovative impact on cutting-edge clean technology. This may be a relatively circularity-poor cluster.

- Cluster 2 – Redox advantage:

Relative to the other two clusters, these countries are putting an increased focus on the two emerging technologies of redox-flow and solid-state batteries. Their patent output related to lead-acid batteries is the lowest of the three clusters and their sodium-ion-related IPF share is close to zero. This cluster contains high-tech industrial nations like the US, Germany, and Taiwan, countries that are known to have explicitly expressed

their ambitions in the field of battery technology. The somewhat less exploratory outline of this cluster does not make it the most potentially pro-circular.

- Cluster 3 – Sodium-ion driven:

540 These countries' focus on sodium-ion and lead-acid batteries, which account for about 35% and 24%, respectively. They have almost no innovative output in solid-state batteries and exhibit a greater share in lithium-sulfur batteries comparing to the other two clusters. This cluster comprises countries like Canada, China, and the UK. The bet on more
545 promising technologies may suggest that there is a high pro-circularity potential to be realised.

Finally, the wordage material available in the patent documents help us to build a more detailed and comprehensive picture of battery development. Trigram analysis indicates that batteries are mutating into more complex com-
550 pacts of technology, able to serve new needs (such as more flexible charging and more mobile applications). There is also some suggestions of pro-circularity as rechargeability and less toxicity seem to be to key organising principles of battery innovation from the outset of our time frame. It is also clear that patents are only a pale indicator of the transition toward a Circular Economy, they allow
555 for a better empirical appraisal of systemic transformation but only imperfectly so. Patents are no silver-bullet, but they are informative and should be kept in the methodological potfolio.

5. Conclusions

Global battery patenting activity has shown a robust upward trend in 2000-
560 2019, which briefly came to a halt in the early 2010s, probably due to the global financial crisis and the subsequent recession. The majority of battery patents originate in Asia while high battery patent intensities are revealed in the performance of several Asian and European countries. From 2000-2009

to 2010-2019 a considerable increase in annual battery patenting activity is
565 observed.

We also found that four battery technologies —redox-flow, solid-state, sodium-ion, and lithium-sulfur batteries— have displayed vibrant growth in recent years. Lithium-ion and other lithium battery technologies have also surged, whilst lead-acid and rechargeable alkaline batteries’ share in battery patenting activity has
570 decreased over the overall timeframe. Three country clusters emerge over the four emerging battery types and the already established lead-acid technology. The first group contains lead-acid-focused countries, another with a higher focus on redox-flow and solid-state batteries, and a third group that contains countries with higher sodium-ion and lithium-sulfur-related patenting shares. A case
575 could be made that these clusters differ in their degree of pro-circularity.

A text mining approach yielded the observation that the battery-related technologies energy storage systems, battery management systems, wireless power transmission, electric vehicle charging, and uncrewed aerial vehicles (i.e., drones) are growing in relevance both in absolute terms and relative to gen-
580 eral battery patenting activity. Generally speaking, there are changes that can be associated to progress toward cleaner, less-toxic, more reusable and more application-adaptable battery solutions. Hints of pro-circularity abound, although not always in a straightforward manner.

To pave the way to a truly circular economy it is essential to consider the
585 place of technologies and organisational arrangements, as well as their interdependencies and complementarities (Silva et al. (2015), De Jesus and Mendonça (2018), Lehmann et al. (2022)). Hence, we have to consider the sources of battery innovation, the rate and direction of technical change, but also assume that storage is part of an evolving socio-technical system (i.e. batteries are no ”silver
590 bullit”). To develop an whole-of-system approach it is necessary to consider the material elements involved in batteries (how scarce they are, how much quantity needed, if they are toxic, etc.) and to go beyond the ”end-of-pipe” mentality so as to encompass their recyclability (the conditions the incorporation of recycled materials and the after-life of batteries in the recycling chain). The intersections

595 between storage and circularity via patenting evidence have only been scratched
on the surface, more work along these lines is surely promising.

Overall, understanding the technological development of "clean tech" through
indicators like patents is always an uncertain task. However, our approach con-
sists of a systematic appraisal of underexploited data and highlights robust re-
600 sults that can be further inquired in the future. In the case of batteries, patent
data are thus found to indicate patterns of progress that are both interesting
(from an analytical perspective) and useful (from a policy perspective). Bat-
teries are a crucial component of a moving circular target as society adapts to
climate crisis.

605 **Data Availability**

The raw data used for this study is available via subscription at PATSTAT
Online. Therefore the third-party data subset generated and used for this work
cannot be made public. Said dataset can be reproduced using PATSTAT Online
and the queries and code made available at the GitHub repository associated
610 with this article, which can be accessed by following this link: https://github.com/ph1001/battery_patents.git.

References

- Aaldering, L.J., Song, C.H., 2019. Tracing the technological development
trajectory in post-lithium-ion battery technologies: A patent-based ap-
615 proach. *Journal of Cleaner Production* 241, 118343. URL: <https://www.sciencedirect.com/science/article/pii/S0959652619332135>,
doi:<https://doi.org/10.1016/j.jclepro.2019.118343>.
- The World Bank, 2022. Labor force, total. <https://data.worldbank.org/indicator/SL.TLF.TOTL.IN>. Accessed: 10. Jan. 2022, 12:07.
- 620 Bird, S., Klein, E., Loper, E., 2009. Natural language processing with Python:
analyzing text with the natural language toolkit. " O'Reilly Media, Inc."

- Caraça, J., Åke Lundvall, B., Mendonça, S., 2009. The changing role of science in the innovation process: From queen to cinderella? *Technological Forecasting and Social Change* 76, 861–867. URL: <https://www.sciencedirect.com/science/article/pii/S0040162508001455>, doi:<https://doi.org/10.1016/j.techfore.2008.08.003>.
625
- Castellaci, F., Grodal, S., Mendonca, S., Wibe, M., 2005. Advances and challenges in innovation studies. *Journal of Economic Issues* 39, 91–121. URL: <https://doi.org/10.1080/00213624.2005.11506782>, doi:10.1080/00213624.2005.11506782, arXiv:<https://doi.org/10.1080/00213624.2005.11506782>.
630
- Dang, H.A.H., Serajuddin, U., 2020. Tracking the sustainable development goals: Emerging measurement challenges and further reflections. *World Development* 127, 104570.
- De Jesus, A., Antunes, P., Santos, R., Mendonça, S., 2018. Eco-innovation in the transition to a circular economy: An analytical literature review. *Journal of cleaner Production* 172, 2999–3018.
635
- De Jesus, A., Antunes, P., Santos, R., Mendonça, S., 2019. Eco-innovation pathways to a circular economy: Envisioning priorities through a delphi approach. *Journal of Cleaner Production* 228, 1494–1513.
640
- De Jesus, A., Lammi, M., Domenech, T., Vanhuyse, F., Mendonça, S., 2021. Eco-innovation diversity in a circular economy: Towards circular innovation studies. *Sustainability* 13, 10974.
- De Jesus, A., Mendonça, S., 2018. Lost in transition? drivers and barriers in the eco-innovation road to the circular economy. *Ecological economics* 145, 75–89.
645
- De Rassenfosse, G., Dernis, H., Boedt, G., 2014. An introduction to the patstat database with example queries. *Australian Economic Review* 47. doi:10.1111/1467-8462.12073.

- 650 Dechezleprêtre, A., Ménière, Y., Mohnen, M., 2017. International patent families: from application strategies to statistical indicators. *Scientometrics* 111, 793–828. URL: <https://doi.org/10.1007/s11192-017-2311-4>, doi:10.1007/s11192-017-2311-4.
- Frietsch, R., Kroll, H., 2020. China’s foreign technology market perspectives. *Deutsch-chinesische Innovationspartnerschaft: Rahmenbedingungen, Chancen und Herausforderungen*, pp. 365–375 Metropolis.
- 655 Griliches, Z., 1990. Patent statistics as economic indicators: A survey. *Journal of Economic Literature* 28, 1661–1707. URL: <http://www.jstor.org/stable/2727442>.
- 660 Harris, C.R., Millman, K.J., van der Walt, S.J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N.J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M.H., Brett, M., Haldane, A., del Río, J.F., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., Oliphant, T.E., 2020. Array programming with NumPy. *Nature* 585, 357–362. URL: <https://doi.org/10.1038/s41586-020-2649-2>, doi:10.1038/s41586-020-2649-2.
- 665 Hsu, D.H., Hsu, P.H., Zhao, Q., 2021. Rich on paper? chinese firms’ academic publications, patents, and market value. *Research Policy* 50, 104319. URL: <https://www.sciencedirect.com/science/article/pii/S0048733321001207>, doi:<https://doi.org/10.1016/j.respol.2021.104319>.
- Hsu, P.H., Lee, D., Tambe, P., Hsu, D.H., 2020. Deep learning, text, and patent valuation. SSRN URL: <https://ssrn.com/abstract=3758388>, doi:<http://dx.doi.org/10.2139/ssrn.3758388>.
- 675 IEA, 2020. Innovation in batteries and electricity storage, a global analysis based on patent data.

- Plotly Technologies Inc., 2015. Collaborative data science. URL: <https://plot.ly>.
- Kittner, N., Lill, F., Kammen, D.M., 2017. Energy storage deployment and innovation for the clean energy transition. *Nature Energy* 2, 17125. URL: <https://doi.org/10.1038/nenergy.2017.125>, doi:10.1038/nenergy.2017.125.
- Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B., Bussonnier, M., Frederic, J., Kelley, K., Hamrick, J., Grout, J., Corlay, S., Ivanov, P., Avila, D., Abdalla, S., Willing, C., 2016. Jupyter notebooks – a publishing format for reproducible computational workflows, in: Loizides, F., Schmidt, B. (Eds.), *Positioning and Power in Academic Publishing: Players, Agents and Agendas*, IOS Press. pp. 87 – 90.
- Kosmadakis, I.E., Elmasides, C., Koulinas, G., Tsagarakis, K.P., 2021. Energy unit cost assessment of six photovoltaic-battery configurations. *Renewable Energy* 173, 24–41.
- Lehmann, C., Cruz-Jesus, F., Oliveira, T., Damásio, B., 2022. Leveraging the circular economy: Investment and innovation as drivers. *Journal of Cleaner Production* , 132146.
- Li, Z., Kuo, Y.K., Mahmud, A.R., Nassani, A.A., Haffar, M., Muda, I., 2022. Integration of renewable energy, environmental policy stringency, and climate technologies in realizing environmental sustainability: Evidence from oecd countries. *Renewable Energy* 196, 1376–1384.
- Losacker, S., 2022. ‘license to green’: Regional patent licensing networks and green technology diffusion in china. *Technological Forecasting and Social Change* 175, 121336.
- Malhotra, A., Zhang, H., Beuse, M., Schmidt, T., 2021. How do new use environments influence a technology’s knowledge trajectory? a patent citation network analysis of lithium-ion battery technology. *Research Policy* 50, 104318. URL: <https://www.sciencedirect.com/science/article/>

- 705 pii/S0048733321001190, doi:<https://doi.org/10.1016/j.respol.2021.104318>.
- Marx, M., Gans, J.S., Hsu, D.H., 2014. Dynamic commercialization strategies for disruptive technologies: Evidence from the speech recognition industry. *Management Science* 60, 3103–3123. URL: <https://doi.org/10.1287/mnsc.2014.2035>, doi:10.1287/mnsc.2014.2035, 710 arXiv:<https://doi.org/10.1287/mnsc.2014.2035>.
- McKinney, W., 2010. Data structures for statistical computing in python, in: Stéfan van der Walt, Jarrod Millman (Eds.), *Proceedings of the 9th Python in Science Conference*, pp. 56 – 61. doi:10.25080/Majora-92bf1922-00a.
- 715 Mendonça, S., Confraria, H., Godinho, M.M., 2021. Appropriating the returns of patent statistics: Take-up and development in the wake of zvi griliches. SSRN doi:<http://dx.doi.org/10.2139/ssrn.3971764>.
- Mendonça, S., Schmoch, U., Neuhäusler, P., 2019. Interplay of patents and trademarks as tools in economic competition. *Springer Handbook of Science and Technology Indicators* doi:https://doi.org/10.1007/978-3-030-02511-3_42. 720 springer Handbooks. Springer, Cham.
- Morales, P., Flikkema, M., Castaldi, C., de Man, A.P., 2022. The effectiveness of appropriation mechanisms for sustainable innovations from small and medium-sized enterprises. *Journal of Cleaner Production* , 133921.
- 725 Neuhäusler, P., Rothengatter, O., Frietsch, R., 2019. Patent Applications - Structures, Trends and Recent Developments 2018. *Studien zum deutschen Innovationssystem 4-2019*. Leibniz Information Centre for Economics - ZBW. Berlin. URL: <http://hdl.handle.net/10419/194274>.
- OECD, 2009. *OECD Patent Statistics Manual*. OECD PUBLICATIONS, 2, 730 rue André-Pascal, 75775 PARIS CEDEX 16.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Pas-

- 735 sos, A., Cournapeau, D., Brucher, M., Perrot, M., Édouard Duchesnay, 2011. Scikit-learn: Machine learning in python. *Journal of Machine Learning Research* 12, 2825–2830. URL: <http://jmlr.org/papers/v12/pedregosa11a.html>.
- Ribeiro, B., Shapira, P., 2020. Private and public values of innovation: A patent analysis of synthetic biology. *Research Policy* 49, 103875.
- 740 Santos, A.B., Bogers, M.L., Norn, M.T., Mendonça, S., 2021. Public policy for open innovation: Opening up to a new domain for research and practice. *Technological Forecasting and Social Change* 169, 120821. URL: <https://www.sciencedirect.com/science/article/pii/S0040162521002535>, doi:<https://doi.org/10.1016/j.techfore.2021.120821>.
- 745 Schmoch, U., Gehrke, B., 2022. China’s technological performance as reflected in patents. *Scientometrics* 127, 299–317. URL: <https://doi.org/10.1007/s11192-021-04193-6>, doi:10.1007/s11192-021-04193-6.
- Sharma, S.S., Manthiram, A., 2020. Towards more environmentally and socially responsible batteries. *Energy & Environmental Science* 13, 4087–4097.
- 750 Silva, J., Oliveira, S., Távora, G., Mendonça, S., et al., 2015. The role of innovation in the future pv and storage markets, in: *Proceedings 31st European Photovoltaic Solar Energy Conference and Exhibition*, pp. 3175–3178.
- Sovacool, B.K., Newell, P., Carley, S., Fanzo, J., 2022. Equity, technological innovation and sustainable behaviour in a low-carbon future. *Nature Human Behaviour* 6, 326–337.
- 755 Stephan, A., Bening, C.R., Schmidt, T.S., Schwarz, M., Hoffmann, V.H., 2019. The role of inter-sectoral knowledge spillovers in technological innovations: The case of lithium-ion batteries. *Technological Forecasting and Social Change* 148, 119718. URL: <https://www.sciencedirect.com/science/article/pii/S0040162517306583>, doi:<https://doi.org/10.1016/j.techfore.2019.119718>.
- 760

- National Statistics; Republic of China (Taiwan), 2022. Labor force, total.
<https://eng.stat.gov.tw/ct.asp?xItem=42761&ctNode=1609&mp=5>. Accessed: 10. Jan. 2022, 12:49.
- Távora, G., Silva, J., Mendonça, S., 2020. Evolução tecnológica no armazenamento de energia: uma perspectiva a partir das patentes de baterias, in: CIES2020-XVII Congresso Ibérico e XIII Congresso Ibero-americano de Energia Solar, LNEG-Laboratório Nacional de Energia e Geologia. pp. 1323–1329.
- Thompson, D.L., Hartley, J.M., Lambert, S.M., Shiref, M., Harper, G.D., Kendrick, E., Anderson, P., Ryder, K.S., Gaines, L., Abbott, A.P., 2020. The importance of design in lithium ion battery recycling—a critical review. *Green Chemistry* 22, 7585–7603.
- Van Rossum, G., Drake, F.L., 2009. Python 3 Reference Manual. CreateSpace, Scotts Valley, CA.
- Vezzini, A., 2014. 15 - lithium-ion battery management, in: Pistoia, G. (Ed.), *Lithium-Ion Batteries*. Elsevier, Amsterdam, pp. 345–360. URL: <https://www.sciencedirect.com/science/article/pii/B9780444595133000157>, doi:<https://doi.org/10.1016/B978-0-444-59513-3.00015-7>.
- Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., et al., 2020. Scipy 1.0: fundamental algorithms for scientific computing in python. *Nature Methods* 17, 261–272. URL: <http://dx.doi.org/10.1038/s41592-019-0686-2>, doi:10.1038/s41592-019-0686-2.
- Wang, S., Zhao, E., Razzaq, H.K., 2022. Dynamic role of renewable energy efficiency, natural resources, and climate technologies in realizing environmental sustainability: Implications for china. *Renewable Energy* 198, 1095–1104.
- Waskom, M., Gelbart, M., Botvinnik, O., Ostblom, J., Hobson, P., Lukauskas, S., Gempertline, D.C., Augspurger, T., Halchenko, Y., Warmenhoven, J., Cole, J.B., de Ruiter, J., Vanderplas, J., Hoyer, S., Pye, C., Miles, A., Swain, C.,

- Meyer, K., Martin, M., Bachant, P., Quintero, E., Kunter, G., Villalba, S.,
 790 Brian, Fitzgerald, C., Evans, C., Williams, M.L., O’Kane, D., Yarkoni, T.,
 Brunner, T., 2021. mwaskom/seaborn: v0.11.2 (august 2021). URL: <https://doi.org/10.5281/zenodo.5205191>, doi:10.5281/zenodo.5205191.
- Winslow, K.M., Laux, S.J., Townsend, T.G., 2018. A review on the growing
 concern and potential management strategies of waste lithium-ion batteries.
 795 Resources, Conservation and Recycling 129, 263–277.
- WIPO, 2022. World intellectual property report 2022 - the direction of innova-
 tion.
- Yildizbasi, A., 2021. Blockchain and renewable energy: Integration challenges
 in circular economy era. Renewable Energy 176, 183–197.
- 800 Zhu, P., Gastol, D., Marshall, J., Sommerville, R., Goodship, V., Kendrick,
 E., 2021. A review of current collectors for lithium-ion batteries. Journal of
 Power Sources 485, 229321.
- Ziegler, M.S., 2021. Evaluating and improving technologies for en-
 ergy storage and backup power. Joule 5, 1925–1927. URL: <https://www.sciencedirect.com/science/article/pii/S2542435121003536>,
 805 doi:<https://doi.org/10.1016/j.joule.2021.07.015>.
- Ziegler, M.S., Trancik, J.E., 2021. Re-examining rates of lithium-ion battery
 technology improvement and cost decline. Energy Environ. Sci. 14, 1635–1651.
 URL: <http://dx.doi.org/10.1039/D0EE02681F>, doi:10.1039/D0EE02681F.

810 Appendices

A.1. Data and Methods

A.1.1. The raw data

This study's foundation is the PATSTAT database De Rassenfosse et al. (2014) provided by the European Patent Office, more precisely the Autumn 2021
815 edition of PATSTAT Online. Transact-SQL or T-SQL is the language used for querying it. The query designed for selecting and downloading the data used for this study is defined in the text file "PATSTAT_Online_query.txt," which is included in the GitHub repository associated with this work, which can be found by following this link:

https://github.com/ph1001/battery_patents.git.

820 The patents that were downloaded from PATSTAT and that make up the raw dataset for this study were all patent applications (including ungranted) that are part of patent families whose intra-family value for the feature "earliest publication year" lies in the time frame of 1999-2019 (the timeframe was later reduced to 2000-2019) and which contain at least one IPC entry matching one
825 of the following codes: **H01M...** (processes or means, e.g., batteries, for the direct conversion of chemical energy into electrical energy), **H02J 3/32** (circuit arrangements for AC mains or AC distribution networks using batteries with converting means), **H02J 7...** (circuit arrangements for charging or depolarising batteries or for supplying loads from batteries), or **B60L 53...** (methods of
830 charging batteries, specially adapted for electric vehicles; charging stations or onboard charging equipment therefor; exchange of energy storage elements in electric vehicles).

PATSTAT Online has the restriction that all SQL queries must begin with a "SELECT" statement. This fact makes analyses of a higher complexity im-
835 possible to achieve inside PATSTAT Online itself. Consequently, data must be queried, downloaded, and then processed in a different environment. The programming language used for all steps after querying the database and downloading the data was Python Van Rossum and Drake (2009) (Version 3.9.7),

more specifically the web application Jupyter NotebookKluyver et al. (2016)
 840 (Version 6.4.3), the data processing libraries pandasMcKinney (2010) (Ver-
 sion 1.3.3) and NumpyHarris et al. (2020) (Version 1.20.3), the visualization
 tools PlotlyPlotly Technologies Inc. (2015) (Version 5.1.0) and SeabornWaskom
 et al. (2021) (Version 0.11.2), the text mining suite Natural Language Toolkit
 (NLTK)Bird et al. (2009) (Version 3.6.5), and the analytics toolboxes Scikit-
 845 learnPedregosa et al. (2011) (Version 0.24.2) and SciPyVirtanen et al. (2020)
 (Version 1.7.1).

Ancillary sources were used. The labor force counts used for scaling were
 downloaded from the World Bank’s websiteThe World Bank (2022) and for the
 specific case of Taiwan from the website of ”National Statistics: Republic of
 850 China (Taiwan).”National Statistics; Republic of China (Taiwan)

A.1.2. Preprocessing and data reduction

Preprocessing and data reduction steps undertaken to obtain the final dataset
 from the raw data downloaded from PATSTAT are defined in the Jupyter Note-
 book ”01_create_dataset.ipynb,” which is included in the GitHub repository
 855 linked above. The following paragraphs contain a summary of these prepro-
 cessing steps.

First, the raw data downloaded from PATSTAT Online was loaded and
 checked for its integrity. Then each patent family’s earliest intra-family values
 for the features ”earliest publication date” and ”earliest publication year” were
 860 determined and added as new columns to every row of the dataset (i.e., they
 were harmonized on patent family level). Like this, patent families can easily
 be assigned to their respective year later during the analyses. Next, all patent
 families were classified and tagged as either ”IPF,” ”singleton,” or ”neither.”
 The resulting tags are stored in the newly created column ”tag.” Next, more
 865 tags for further data selection were created. This process took place in five steps
 as described below:

- First, every patent family was scanned for the IPC codes related to non-
 active battery parts, electrodes, or secondary cells (IPC codes H01M 2...,

H01M 50..., H01M 4..., and H01M 10...). Patent families containing any
870 of these codes were added in their entirety, except if they contained any
of the IPC codes H01M 6..., H01M 8..., H01M 12..., H01M 14..., or H01M
16..., which are related to primary cells, fuel cells, hybrid cells, electro-
chemical current or voltage generators not provided for in groups H01M
6/00-H01M 12/00, and structural combinations of different types of elec-
875 trochemical generators, which were hereby explicitly excluded from the
analysis. The patent families passing this stage were tagged as "non-active
parts, electrodes, secondary cells."

- In a second step, every patent family was scanned for the IPC codes related
to "circuit arrangements for ac mains or ac distribution networks using
880 batteries with converting means" (H02J 3/32), "circuit arrangements for
charging or depolarising batteries or for supplying loads from batteries"
(H02J 7...), "methods of charging batteries, specially adapted for elec-
tric vehicles" (B60L 53...), or "secondary cells; methods for charging or
discharging" (H01M 10/44). Patent families that contained any of these
885 codes were added in their entirety, except if they contained any of the
IPC codes listed for exception in the above step or any of the codes B60L
53/54, B60L 53/55, or B60L 53/56 that refer to charging stations using
fuel cells, capacitors, or mechanical storage means, respectively. Patent
families that passed this stage were tagged as "charging."

- 890 • As a third step, to identify affiliations of the resulting patent families to
a set of technological categories, each patent family's titles and abstracts
were scanned using individual sets of regular expressions for each tech-
nology. These regular expressions are defined in the Jupyter notebook
"01_create_dataset.ipynb." Titles and abstracts of all languages were con-
895 sidered and a patent family was selected in its entirety if any substring of
its titles or abstracts matched any of the respective regular expressions.
Note that—to decrease the risk of false positives—before scanning ab-
stracts for these regular expressions, they were cut off at the beginning of

any appearance of the string "independent claims are also included for."

900 The selected patent families were assigned the value 1 in the newly created columns with the column name "is x," with $x \in \{\text{Lead-acid, Lithium-air, Lithium-ion, Lithium-sulfur, Other Lithium, Magnesium-ion, nickel-cadmium, nickel-iron, nickel-zinc, nickel-metal hydride, Rechargeable alkaline, Sodium-sulfur, Sodium-ion, Solid-state, Aluminium-ion, Calcium-ion},$
905 ion), Organic radical} being the name of the respective technology. Please note that due to the considerable overlap of the concept of solid-state batteries with other technologies, especially lithium-ion batteries, all patent families that were classified as patents related to solid-state batteries were untagged in any other category in which they acquired tags through the
910 process described here. To be very clear: This especially means that the lithium-ion battery category does not contain any patent families tagged as solid-state battery inventions.

- The fourth step's purpose was to add patent data related to redox-flow and nickel-hydrogen batteries to the dataset. For this purpose, a combination of IPC classes queries and text queries was deployed. The reason
915 for this separate step is that redox-flow and nickel-hydrogen batteries are closely related to fuel cells. Consequently, patents associated with them are often included in IPC classes that were excluded by the above steps. Analogous to the above steps, the IPC classes qualifying for potential inclusion were H01M 2..., H01M 50..., H01M 4..., H01M 8..., and H01M
920 10... and the IPC classes demanding exclusion were H01M 6..., H01M 12..., H01M 14..., and H01M 16.... Analogous to the above step, these patent families' titles and abstracts were then scanned using one set of regular expressions for redox-flow and another for nickel-hydrogen batteries. These regular expressions can be reviewed in the Jupyter notebook
925 "01_create_dataset.ipynb." All patent families that passed this stage were assigned the value 1 in the newly created columns with the names "is redox-flow" or "is nickel-hydrogen," respectively.

- As the last step, another additional column was computed: The dataset
930 column "technologies one hot sum" contains the sum across each row's "is
<technology name>" values. This sum is needed in the rare cases where
technology classifications overlap. The share of patent families with more
than one technology associated with them was 0.61% in the final dataset.
The counts resulting from these overlapping technologies were not counted
935 multiple times but, using the respective "technologies one hot sum" value,
distributed as equal fractions across the overlapping classes.

The tags created in the above steps were used for selecting the appropriate
data for each analysis. All patent families not having the "IPF" tag were fil-
tered out before all analyses. They were kept in the unfiltered dataset only for
940 completeness, having potential future analyses with a broader scope in mind.
The data selection method applied before each analysis that is based on the
labels whose creation was described above is presented in Fig. A.9:

A.1.3. Counting patents

As already mentioned in the Introduction, the methodological setup of this
945 study roughly follows the framework defined in the IEA and EPO report IEA
(2020). This means that all dates in this study refer to the earliest publication
date within the respective IPF, and the geographic distributions were calcu-
lated based on the geographic information assigned to the respective inventors
in PATSTAT. Each inventor was assigned an equal fraction of the respective
950 count where multiple inventors were indicated. We believe there is a limitation
to this approach, which is described as follows: For identifying the inventors,
their PATSTAT name attribute "psn_name" is used. The harmonization of this
feature, which PATSTAT carried out, is not complete. For example, pairs of
entries like "KERUEL BERNARD" and "BERNARD KERUEL" exist, which
955 in reality correspond to the same inventor, but are consequently treated as two
different individuals. This shifts the fractions of countries of origin in these
entries' patent families in favor of the country of the unharmonized name.

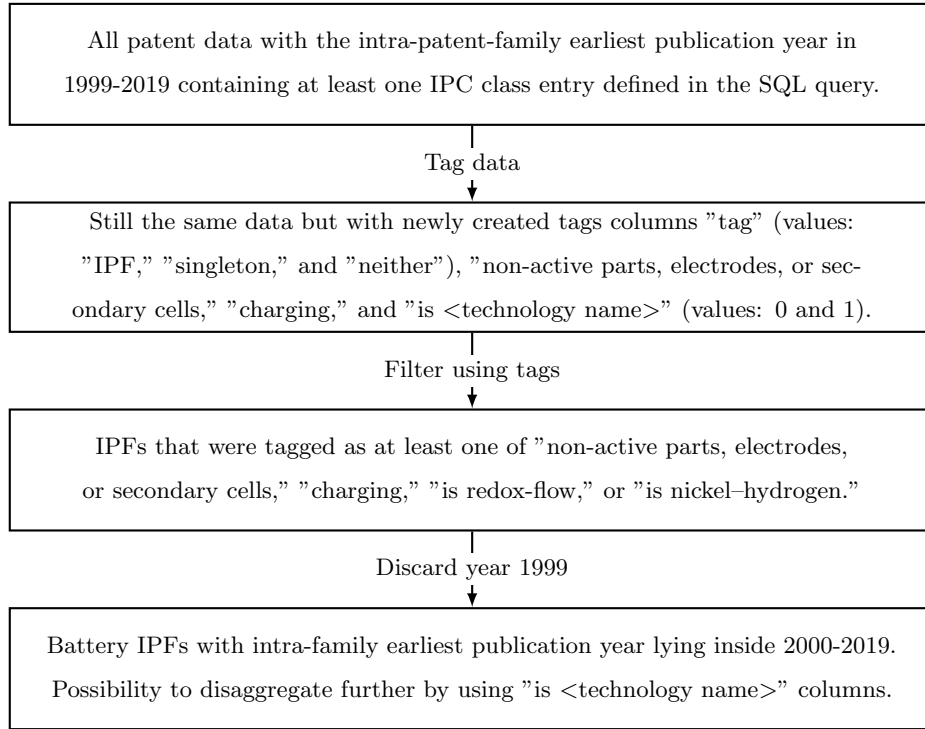


Figure A.9: Flow chart depicting the data selection process for this study. The entire raw dataset was labeled using newly created columns. Before each analysis, the final dataset was acquired by filtering, using labels and timestamp columns.

The code used for counting patents by countries is contained in the Jupyter Notebook "02_counts_technologies_clustering.ipynb," which is part of the GitHub repository linked at the beginning of this section.

A.1.4. Methods: Battery technologies

Unlike the IEA and EPO report IEA (2020), in the current study, fractional counting was also applied when breaking down counts by technological categories. Whenever an IPF was classified as belonging to more than one category, each technology was assigned an equal fraction of the respective count. This situation only happened in a tiny minority of the cases since only 0.61% of all IPFs were assigned to more than one technology. The code used for counting patents by technologies is contained in the Jupyter Notebook

"02_counts_technologies_clustering.ipynb," which is part of the GitHub repository linked at the beginning of this section.

A.1.5. Methods: Clustering

The metric R^2 applied for comparing the performance of several clustering algorithms using varying numbers of clusters can be characterized as follows:

$$R^2 = \frac{SSB}{SST} = \frac{SST - SSW}{SST} = 1 - \frac{SSW}{SST} \in [0, 1] \quad (\text{A.1})$$

where $SSB = \sum_{i=1}^p n_i (\bar{X}_i - \bar{X})^2$ = *sum of squared differences between groups*
 975 and $SSW = \sum_{i=1}^p \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2$ = *sum of squared differences within groups*
 and $SST = \sum_{i=1}^p \sum_{j=1}^{n_i} (X_{ij} - \bar{X})^2$ = *total sum of squared differences* with
 p = number of clusters, n_i = number of elements in cluster i , \bar{X}_i =
centroid of cluster i , \bar{X} = *center of whole dataset*, and X_{ij} = *j th element of cluster i* .
 Relations A.1 are true if and only if $SST = SSW + SSB$, which is the case
 980 because:

$$\begin{aligned} SST &= \sum_{i=1}^p \sum_{j=1}^{n_i} (X_{ij} - \bar{X})^2 = \sum_{i=1}^p \sum_{j=1}^{n_i} (X_{ij} - \underbrace{\bar{X}_i + \bar{X}_i}_{=0} - \bar{X})^2 = \sum_{i=1}^p \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i - \bar{X})^2 \\ &= \sum_{i=1}^p \sum_{j=1}^{n_i} (\bar{X}_i - \bar{X})^2 + 2 \sum_{i=1}^p \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)(\bar{X}_i - \bar{X}) \\ &= \sum_{i=1}^p \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2 + \sum_{i=1}^p \sum_{j=1}^{n_i} (\bar{X}_i - \bar{X})^2 + 2 \sum_{i=1}^p (\bar{X}_i - \bar{X}) \underbrace{\sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)}_{=0} \\ &= \sum_{i=1}^p \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2 + \sum_{i=1}^p \sum_{j=1}^{n_i} (\bar{X}_i - \bar{X})^2 \\ &= \underbrace{\sum_{i=1}^p \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2}_{(\text{A.1.5})} + \underbrace{\sum_{i=1}^p n_i (\bar{X}_i - \bar{X})^2}_{(\text{A.1.5})} = SSW + SSB \end{aligned}$$

with

$$\sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i) = \sum_{j=1}^{n_i} X_{ij} - \sum_{j=1}^{n_i} \bar{X}_i = \frac{n_i}{n_i} \sum_{j=1}^{n_i} X_{ij} - n_i \bar{X}_i = n_i \bar{X}_i - n_i \bar{X}_i = 0 \quad (\text{A.2})$$

A higher R^2 value indicates a better clustering solution, given a non-varying dataset and a fixed number of clusters. Clustering algorithms that were compared are k-means and hierarchical agglomerative clustering using complete,
 990 average, single, and Ward linkage. The numbers of tested clusters ranged from two to nine.

The decision to use only the five dimensions "lead-acid," "redox-flow," "solid-state," "sodium-ion," and "lithium-sulfur" resulted from extensive testing of other configurations, especially those that included "Lithium-ion," "Other lithium,"
 995 or a joint category of "Lithium-ion and other lithium." These tests were not found to be satisfying since it was observed that the lithium-related IPFs were overshadowing the other categories due to their sheer amount, resulting in clustering solutions that lacked the clear interpretability of the solution presented in this work. Lithium-air batteries, another battery technology that
 1000 has received increased attention in recent years, Aaldering and Song (2019) was considered a candidate feature for this analysis but was discarded due to its still very low yearly IPF counts. The code used for clustering countries based on their technology distribution is contained in the Jupyter Notebook "02_counts_technologies_clustering.ipynb," which is part of the GitHub repository
 1005 linked at the beginning of this section.

A.1.6. *Methods: Title and abstract mining*

Unigrams, bigrams, and trigrams were extracted from cleaned abstracts and titles from which meaningless words and phrases had been removed and in which certain synonyms and anomalies had been treated. The n-gram counts method
 1010 simply counts occurrences and displays them as annual sums. In contrast, the n-gram intensities method does the same with the difference that its resulting values are scaled using each years' numbers of abstracts or titles, respectively. Three ways for presenting the identified n-grams were designed for this study:

- Method 1a: Sorted in descending order of increase over the given time-
 1015 frame of 2000-2019 with the measure used for sorting being $m_1 = count_{last} - count_{first}$.

- Method 1b: Sorted in ascending order of increase over the given timeframe of 2000-2019 with the measure used for sorting being m_1 . This method's purpose is to show n-grams that exhibit a negative increase, i.e., have decreased over the given time period.

- Method 2: Sorted in descending order with the measure used for sorting being $m_2 = \sum abs(year - to - year\ difference_{i,i+1})$. This method's purpose is to show n-grams whose count or intensity changed the most (in absolute terms) between all adjacent years.

The results displayed in the tables that are presented in this study were obtained using method 1a, patent abstracts, and trigrams. The code for computing these results is contained in the Jupyter Notebook "03_title_and_abstract_mining.ipynb," which is part of the GitHub repository linked at the beginning of this section. The results obtained by using the methods and data combinations not presented in this paper can best be viewed by opening the HTML file "03_title_and_abstract_mining.html," which is also available in the same folder. The combinations for which results were computed can be characterized by the Cartesian product $c = \{n = 1, n = 2, n = 3\} \times \{n\text{-gram counts}, n\text{-gram intensities}\} \times \{method\ 1a, method\ 1b, method\ 2\} \times \{titles, abstracts\}$.

A.2. Battery IPF intensities for each continent

Figure B.10 presents the development of the number of battery IPFs per 1M workers (battery IPF intensities) for each continent. In terms of battery IPF intensities, Europe and North America outperform Asia. Asia contributed approximately 60% to the global labor force in the 2000-2019 timeframe (Europe and North America contributed approximately 9% and 8%, respectively). This imbalance explains why Asia's battery patenting activity is lower in the perspective of this representation.

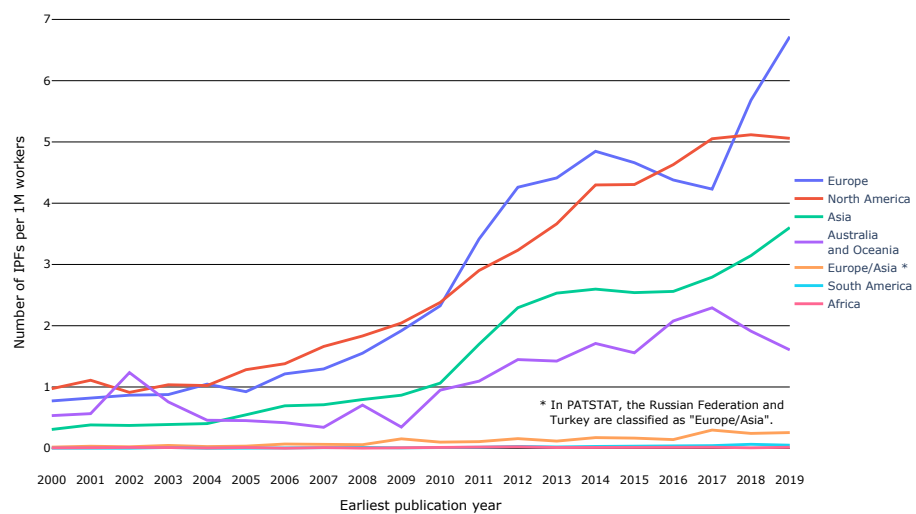


Figure B.10: Battery IPFs per 1M workers by inventors' continents of origin, 2000-2019. In terms of battery IPF intensities, Europe and North America outperform Asia.