A Bibliometric based Analysis to Identify Promising Domains of Decarbonisation Technologies

Paulo Moisés Costa
Electrical Department
Polytechnic Institute of Viseu
Viseu, Portugal
paulomoises@estgv.ipv.pt

Paulo Tomé
Computer Science
Polytechnic Institute of Viseu
Viseu, Portugal
ptome@estgv.ipv.pt

Bruno Almeida
Electrical Department
Polytechnic Institute of Viseu
Viseu, Portugal
estgv16783@alunos.estgv.ipv.pt

Nuno Bento
Instituto Universitario de Lisboa (ISCTE-IUL)
DINAMIA'CET
Lisboa, Portugal
nuno.bento@iscte-iul.pt

***Abstract*—The high volume of research outputs related to technologies and strategies for decarbonisation can make it challenging to understand the relevance of the presented proposals. In this paper, a bibliometric-based analysis is proposed to understand better the domains and subdomains of decarbonisation technologies that have given more attention by the scientific community . The results show that carbon capture technologies, renewable energy sources, electric mobility, green hydrogen, and storage systems are the technology innovations that have received the most attention in the period 2013 to 2021. These results confirm the traditional supply-side orientation of research on technology innovations.**

***Index Terms*—Decarbonisation, bibliometric analysis, data mining, energy system.**

I. INTRODUCTION

Action against climate change must accelerate soon if a major disruption is to be avoided [1]. The evolution in climate change highlights the need to take measures for decarbonisation. In fact, climate change is one of the biggest challenges that humanity has to face today. The main drivers of climate change are generation, transmission, transformation, and energy use (essential to modern life). A significant part of the energy consumed in the activities of modern societies comes from fossil sources, resulting in a massive release of greenhouse gases (such as methane, carbon dioxide, and nitrous oxide), responsible for the phenomenon of global warming and its consequences [2]. Furthermore, the rapid growth of the worldwide population and the increase in their standard of living resulted in a rapid rise in energy demand, intensifying the use of fossil fuels over the last decades [2]. Besides the climatic challenges, the intensive usage of fossil energy has other non desired consequences, namely concerning people's health. Therefore, modern society is at a critical crossroads, looking for alternatives that allow decarbonisation while seeking to minimise the impacts on the economies that may result from it. In the last years, the scientific community has been developing intensive research and evaluation of potential technologies and strategies to mitigate the impact of the energy sector on the global warming problem. Several approaches have been proposed, namely, related to improving the efficiency of conversion [3] and energy use [4], [5], the increasing use of reduced or zero-emission technologies to generate energy [6], or technologies for carbon capture [7]. Still, previous research indicate that the direction of search often favours large-scale, supply-side energy technologies, overlooking more granular approaches namely for the energy demand [8], [9]. The high volume of research related to decarbonisation technologies and strategies can make it challenging to understand the relevance of the proposals presented.

In this work, two readily available text mining software are used to identify the technology innovations that may be the most promising for the decarbonisation process. Concretely, the European Commission's TIM ("Tools for Innovation Monitoring") [10] and the VOSviewer from Leiden University [11] are used to identify technological trends in literature, projects, and patents. In the scientific literature, some works look already for particular technology fields. For instance, in [12], the literature about energy storage and its importance in achieving decarbonisation is analysed. The authors use a bibliometric analysis for this purpose. This analysis is also used in [13] to assess the impact of the diffusion of energy efficiency technologies on the decarbonisation of the European housing stock. The authors of [14] present a bibliometric perspective on developing Carbon Capture Utilization and Storage research in China. In [15], the authors aim to characterise and map the body of knowledge on biomass and organic waste potentials toward implementing circular bioeconomy

The authors acknowledge the financial support from FCT – Foundation for Science and Technology, I.P., to the research project Sus2Trans (PTDC/GES-AMB/0934/2020). Paulo Costa and Paulo Tomé also acknowledges funding from FCT of the project Refª UIDB/05583/2020 and thanks the Research Centre in Digital Services (CISeD) and the Polytechnic of Viseu for their support.

platforms. The existing literature focuses on the evaluation of specific technologies or measures. Indeed, as far as the authors know, a more comprehensive study of the most promising technologies for the decarbonisation process is still missing. The main contribution of this article is to fill this gap. Moreover, the existing literature usually presents bibliometric analysis supported by just one tool. In this work, two text mining tools are simultaneously used to include several research outputs (papers, patents) and improve confidence in the results.

II. Methodology

Two bibliometric analysis software were used, namely the TIM and the VOSviewer. The TIM software tracks established and emerging technologies by retrieving bibliometric data from various sources, namely the SCOPUS, CORDIS, and PATSTAT databases [16]. The VOSviewer is a software that allows performing different types of analysis according to a set of data, which must be previously extracted from one of the following databases: Web of Science, Scopus, PubMed, RIS, or Crossref JSON. In both cases (TIM and VOSviewer), the dataset extracted is defined using a boolean search string, which must be designed to capture the desired information. Therefore, the search string must be composed of a suitable set of terms.

The methodology used in this work is outlined in Fig. 1. The various steps are explained in the following subsections.

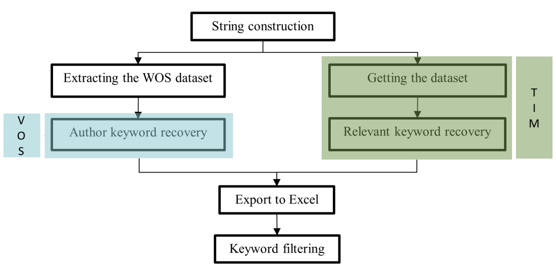

Fig. 1. Schematic of the methodology used

A. Search String Design

The search string was designed based on a previous literature review about energy sector decarbonisation, which allows defining a set of decarbonisation-related terms. We chose to use a search string with two parts linked through a logical AND, as can be seen in Algorithm 1.

This approach forces each document in the retrieved dataset to contain at least one of the terms of both string parts, minimising the retrieval of irrelevant data and avoiding exceeding the limit of 10,000 documents that can be treated for the TIM software. Moreover, the string was designed to capture the most relevant domains of technologies being studied to decarbonise the energy sector.

B. VOSviewer analysis

The search string was used to extract the dataset from the Web of Science database to be subsequently analysed through

Algorithm 1 String design

(("Transformation pathways" OR "co2 emission" OR "greenhouse gas emission" OR "technological innovations" OR "2050" OR "System transformation" OR "2030" OR "global warming" OR "climate solution" OR "climate target" OR "climate policy" OR "displace fossil fuels" OR "1.5º" OR "ghg emission" OR "green house gas" OR "Paris Agreement" OR "transition in electricity" OR "energ* transition" OR "clean energ*" OR "sustainable energy" OR "new energ*" OR "carbon emission" OR "climate change" OR "mitigation" OR "technology" OR "disruptive") AND ("decarbonization" OR "carbon reduction" OR "low carbon" OR "emission reduction" OR "zero carbon" OR "decarbonised"))

VOSviewer software, as shown in Fig. 2. The dataset was extracted in the "Tab Delimited File" format and with the content "Full Record and Cited References" required for using the dataset in the VOSviewer software. The use of the Topic option imposes that the search is carried out in the title, abstract, and keywords of the documents included on the Web of Science database.

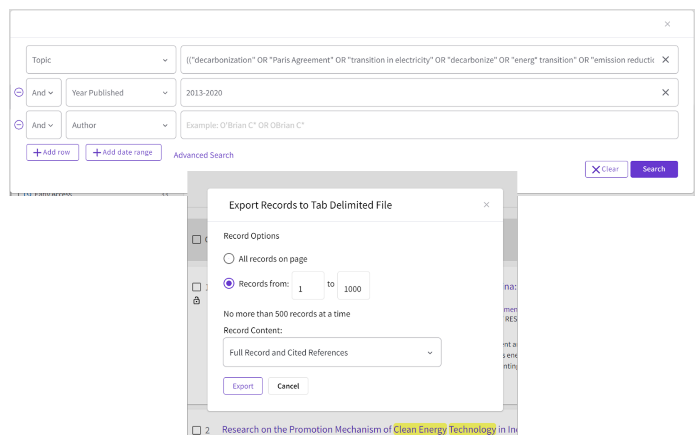

Fig. 2. Dataset extraction from the Web Of Science database

The dataset is then used as VOSviewer's input to obtain the desired information, the "author keywords" of the documents, and their occurrence. So the "co-occurrence" option was adopted, as shown in Fig. 3.

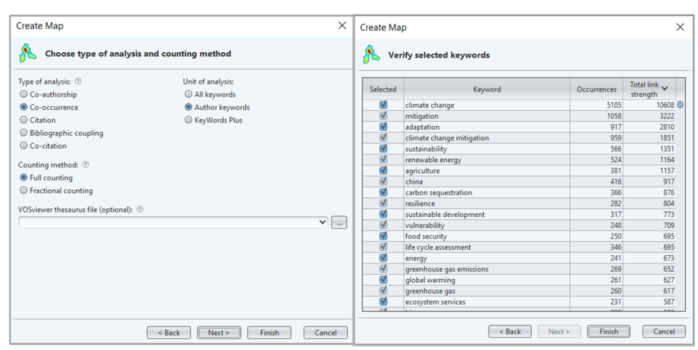

Fig. 3. Co-occurrence analysis in VOSviewer

C. TIM analysis

In the TIM platform, the procedure differs slightly, given that this software does not impose a previous dataset extraction. TIM automatically retrieves the dataset that respects the condition defined by the search string. As shown in Fig. 4, the search string was preceded by the term ti_abs_key, which imposes that the search is carried out in the title, abstract, and keywords of the documents included on the source databases.

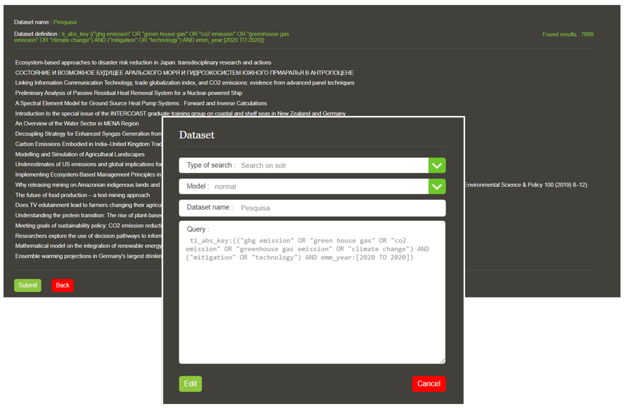

Fig. 4. TIM Dataset definition

After obtaining the dataset, TIM may be used to classify the keywords according to different algorithms. In this work, the ”Relevant Keywords” algorithm was used. This algorithm ranks the keywords by a relevance value defined by a modified version of the classic term Frequency-Inverse Document Frequency (TF-IDF). TF-IDF assigns different weights to keywords according to where they appear in the document, namely 1 when the keywords are in the document's title, 0.5 when in the abstract, and 2 when in the keyword field [17]. Therefore, TIM's meaning of ”relevance” should not be directly compared to the ”occurrence” obtained through VOSviewer.

Keyword	Relevance
low carbon (LC)	338
carbon emissions (CE)	325
emission reduction (ER)	203
co2 emission (CE)	156
de carbonization	151
...	
carbon capture storage (...	60
green house gas (GHG)	58
carbon tax (CT)	55
carbon price (CP)	50
energy transmitter (ET)	50
china s	48
...	
renewable energy (RE)	42
carbon footprint (CF)	42

Fig. 5. TIM's ranked list of Relevant Keywords list

D. Keyword treatment

Once the keywords related to the defined search string are obtained, it is necessary to process them. Indeed, the list of keywords found in both TIM and VOSviewer includes many concepts/words irrelevant to the intended study (for instance, in Fig. 5, the term emission reduction is not a decarbonisation technology domain). Thus, the obtained keywords need to be filtered to eliminate the undesired terms/words. A custom semantic dictionary was created to do the filtering process. This dictionary aggregates various versions of the same meaning (e.g. PV system, photovoltaic, solar PV, etc.) and allows not to consider terms and words irrelevant to the desired analysis (e.g. low carbon, carbon emissions, etc.). The dictionary's construction implied successive iterations to maximise the aggregation of terms/words assumed as synonymous. Fig. 6 shows an extract of the created dictionary used with an algorithm implemented in an Excel VBA program to filter the keywords list obtained through TIM and VOSviewer.

	A	B	C	D
1	renewable energy source	renewable energy	renewable source	variable renewables
2	renewable electricity	renewable electricity (re)	renewable power	
3	solar energy	solar panel	solar power	solar energy (se)
4	photo voltaic (pv)	photovoltaic	solar photovoltaic	solar pv
5	solar termal	solar thermal (st)	concentrated solar thermal	
6	wind power	wind energy	wind plant	wind power (wp)
7	onshore wind			
8	horizontal axis wind turbine (hawt)			
9	offshore wind	offshore wind (ow)	floating offshore wind turbi	offshore renewable

Fig. 6. Extract of the semantic dictionary

Fig. 8 evidences the procedure that is performed by the implemented algorithm. In practice, the algorithm reads the keywords from the list returned by TIM or VOSviewer, previously exported to Excel format (as shown in Fig. 7), and checks whether or not there is a match in the dictionary. When a correspondence exists, the value of relevance/occurrence of the keyword is added to the already existent value. The unmatched keywords are stored to a specific Excel sheet. Therefore, the developed algorithm group the keywords with similar meanings, thus obtaining a clearer and more realistic view of each keyword's occurrence/relevance.

	A	B
1	Keyword	Relevance
2	low carbon (LC)	338,2883
3	carbon emissions (CE)	325,5975
4	emission reduction (ER)	203,5281
5	co2 emission (CE)	156,1575
6	de carbonization	151,1592
...		
14	carbon capture storage (CCS)	60,75693
15	green house gas (GHG)	58,03194
16	carbon tax (CT)	55,41615
17	carbon price (CP)	50,64754
18	energy transmitter (ET)	50,38329
19	china s	48,45521
...		
26	renewable energy (RE)	42,57052
27	carbon footprint (CF)	42,3035

Fig. 7. Exported keywords list

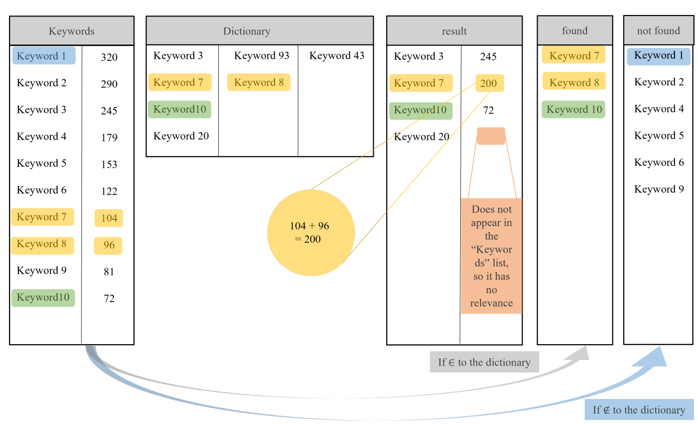

Fig. 8. Schematic of how the algorithm works

TABLE I
TIM RESULTS

TIM	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	**2020**	**Total**
carbon capture storage	290	287	180	167	343	252	276	205	2000
renewable energy source	84	102	114	128	135	142	180	147	1031
photo voltaic	35	60	56	65	93	67	98	72	547
electric vehicle	48	37	47	60	64	68	88	75	488
energy transmitter	19	20	26	39	50	74	85	66	379
net zero energy building	24	38	34	39	42	45	51	80	353
green hydrogen	15	12	28	27	36	42	67	71	298
wind power	34	31	36	36	46	33	43	30	287
solar energy	26	37	38	33	37	44	39	33	285
battery	12	22	20	22	45	39	55	42	256
nuclear power	43	29	30	39	28	25	33	23	251
storage	16	20	21	25	32	39	38	33	224
energy efficiency	38	24	29	29	30	28	27	19	223
bio fuel	26	28	25	21	31	34	23	23	210
energy mix	18	13	22	18	26	31	42	41	209
wind and solar	14	20	19	22	28	28	36	30	196
heat pump	18	18	21	21	31	28	35	21	193
biomass	20	23	23	21	17	30	34	22	191
distributed generation	18	24	15	28	28	25	24	25	186
combined heat power	26	20	28	25	27	26	21	11	184
...									

TABLE II
VOS RESULTS

VOS	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	**2020**	**Total**
renewable energy source	33	55	64	78	111	115	181	245	882
carbon capture storage	42	62	50	39	64	55	61	69	442
energy efficiency	22	29	33	39	38	60	59	56	336
electric vehicle	9	10	19	14	25	26	32	48	183
wind power	13	10	12	17	15	28	28	34	157
photo voltaic	8	15	12	14	23	16	27	36	151
green hydrogen	9	7	5	9	10	16	29	50	135
biomass	11	13	13	13	18	19	19	27	133
solar energy	7	7	9	9	20	13	20	31	116
storage	5	6	5	6	18	14	20	39	113
bio energy	3	7	12	14	12	20	10	22	100
bio fuel	5	4	7	13	10	10	11	13	73
net zero energy building	4	15	7	9	11	9	5	13	73
smart grid	6	9	9	8	13	5	9	9	68
nuclear power	7	9	5	6	15	6	5	9	62
battery	6	2	5	4	9	5	14	13	58
heat pump	4	4	8	6	10	7	10	8	57
distributed generation	4	4	5	3	10	9	10	11	56
circular economy	4	1	1	1	6	5	19	18	55
demand response	2	0	8	5	10	10	10	6	51
...									

III. APPLICATION

The methodology described in section II was used to explore relevant databases of publications, patents and projects to identify domains of technologies pointed out by the scientific community as the more suitable for the decarbonisation of the energy sector. Fig. 9 shows, for the 2013 to 2020 period, the evolution of the number of references retrieved by TIM and VOSviewer software. As expected, the number of references increased, except for 2020, probably due to the beginning of the Covid-19 pandemic. This increasing number of references evidences the scientific community's interest in the topic of decarbonisation.

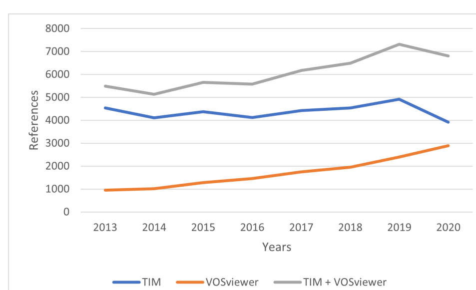

Fig. 9. Number of documents retrieved by the TIM and VOS

It is importat to mention, for the case of TIM software, the retrieved references are not just scientific papers but also research projects and patents registration, as shown in Figure 10.

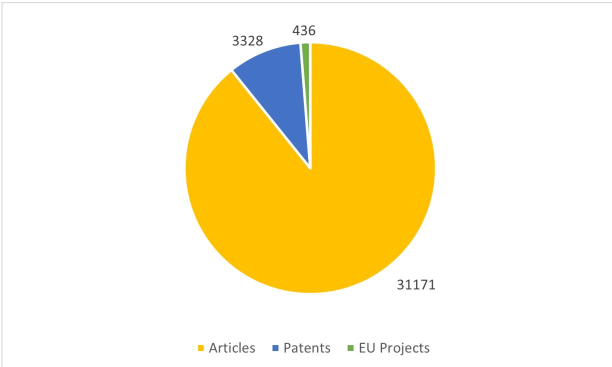

Fig. 10. Documents retrieved by the TIM

Table I and table II show extracts of the research results in TIM and VOSviewer, namely the first 25 terms obtained after applying the previously described filtering process. It is essential to clarify that there may be some overlap in the values of relevance/occurrence obtained for several of the terms in table I and table II. In other words, documents that refer to a specific term may also quote another term present in the same table. For example, the results for the "Renewable Energie Source" term may include contributions from documents that also are accounted for in terms such as "solar energy", "wind power", "photovoltaic", etc. (and vice-versa). Moreover, the terms in the tables tend to not coincide with specific technologies but rather with domains of technologies suitable for decarbonisation, which is the objective of this article. For example, the Renewable Energy domain may include multiple specific technologies (such as Floating Offshore Wind or Perovskite Solar Cell), which may not be present in table I and table II.

Table III was built based on the information presented in table I and table II by ranking the terms based on the sum of the respective values of the number of occurrences (VOSviewer) with the relevance (TIM). Observing the ordered list obtained, we can define a set of domains (highlighted in blue) and subdomains (highlighted in yellow) of technologies associated with decarbonisation that have been deserving greater attention from the scientific community in recent years. Therefore, the bibliometric analysis software allows us to overview the technologies that the scientific community considers to promote decarbonisation. For instance, it is possible to perceive that carbon capture and storage, renewable energy generation, electric mobility, and efficiency in energy use are the domains with significant preponderance in the scientific community's work. Fig. 11 shows a glimpse of an hypothetical future energy system if the relevant domains identified emerge and become prevalent.

TABLE III
RESULTS

Keywords	TIM	VOS	TOTAL
carbon capture storage	2000	442	2442
renewable energy source	1031	882	1913
photo voltaic	547	151	698
electric vehicle	488	183	671
energy efficiency	223	336	559
wind power	287	157	444
green hydrogen	298	135	433
zero energy building	353	73	426
energy transmitter	379	8	387
storage	224	113	337
biomass	191	133	324
nuclear power	251	62	313
bio fuel	210	73	283
bio energy	179	100	279
smart grid	119	68	187
district heating	144	38	182
fuel cell	106	32	138
hydropower	103	30	133
demand response	74	51	125
thermal energy system	109	15	124
microgrids	68	32	100
ammonia	80	19	99
smart city	57	37	95
geothermal energy	51	20	71
neg. emission technologies	29	7	36

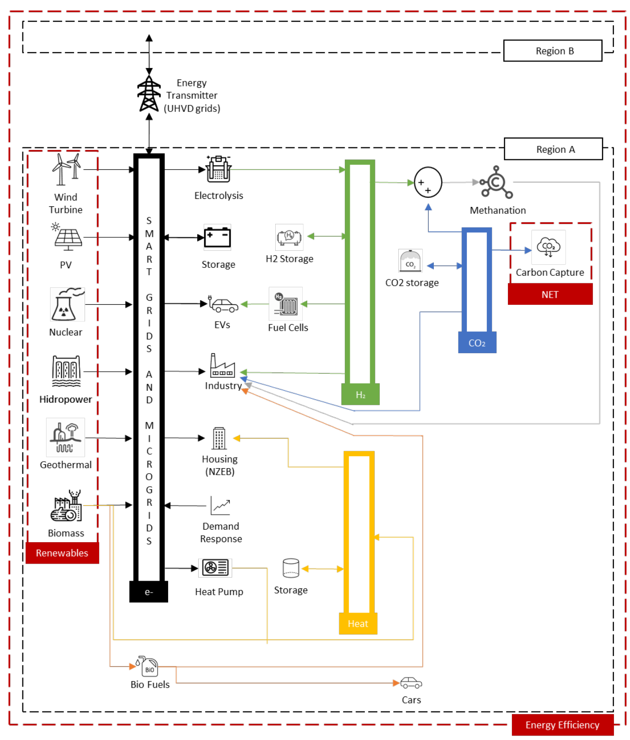

Fig. 11. Relevant domains of technologies

IV. CONCLUSION

The evolution in climate change highlights the need to decarbonise the economy in general and, in particular, the energy sector. The scientific community have been making significant efforts to develop technologies and strategies that allow decarbonisation. As a result of these efforts, many documents have been published, research projects developed, and patents registered. The high volume of research makes it challenging to understand the relevance of the presented proposals. Therefore, a methodology based on the use of two text mining tools is proposed in this paper to perceive the domains of technologies that may be more promising for decarbonisation. The application of the proposed methodology for the period 2013-2020 revealed an increasing number of documents related to decarbonisation technologies. The exception is 2020 (probably due to the onset of the pandemic). Therefore, we may conclude that the scientific community has a growing interest in decarbonisation.

Moreover, the results evidence a comprehensive set of promising technologies for the decarbonisation process. The carbon capture technologies, the renewable energy sources, the electric mobility, the green hydrogen, and the storage systems are the domains of technologies that have received the most attention from the scientific community in the period 2013-2020. It is also important to highlight that the promising domains of technologies may be framed into three fundamental groups: the energy generation side (e.g. electricity and hydrogen generation), the energy demand side (e.g. electric mobility, demand response), and the infrastructure side (e.g. smartgrids, microgrids). The results point to an excessive emphasis on large technologies on the supply side and that more distributed approaches, namely on the demand side with great potential for decarbonisation [9], maybe being overlooked by research. The reliance on literature, projects and patents can explain this potential analysis bias. Therefore we will proceed with additional research to identify potentially disruptive technologies for decarbonisation. The subsequent step includes characterising technologies across dimensions such as their technological readiness, associated costs, modularity, and the potential impact on the decarbonisation process. Despite all care taken in defining the search string in the bibliometric analysis, there are always limitations in the proposed methodology, as there may be a potential influence on the results of the chosen string. Similarly, the results may also be influenced by the databases used. Nevertheless, this limitation is mitigated by using two different bibliometric analysis tools, which use different databases.

REFERENCES

[1] IPCC, "Climate Change 2022: Mitigation of Climate Change. Contribution of Working Group III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change," J. M. P.R. Shukla, J. Skea, R. Slade, A. Al Khourdajie, R. van Diemen, D. McCollum, M. Pathak, S. Some, P. Vyas, R. Fradera, M. Belkacemi, A. Hasija, G. Lisboa, S. Luz, Ed. Cambridge University Press, Cambridge, UK and New York, NY, USA, 2022, doi:10.1017/9781009157926.

[2] A. G. Olabi and M. A. Abdelkareem, "Renewable energy and climate change," *Renewable and Sustainable Energy Reviews*, vol. 158, p. 112111, Apr. 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1364032122000405

[3] M. Borasio and S. Moret, "Deep decarbonisation of regional energy systems: A novel modelling approach and its application to the Italian energy transition," *Renewable and Sustainable Energy Reviews*, vol. 153, p. 111730, jan 2022. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1364032121010030

[4] M. D. Obrist, R. Kannan, T. J. Schmidt, and T. Kober, "Long-term energy efficiency and decarbonization trajectories for the Swiss pulp and paper industry," *Sustainable Energy Technologies and Assessments*, vol. 52, Aug. 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2213138821009516

[5] M. Andrei, P. Thollander, I. Pierre, B. Gindroz, and P. Rohdin, "Decarbonization of industry: Guidelines towards a harmonized energy efficiency policy program impact evaluation methodology," *Energy Reports*, vol. 7, pp. 1385–1395, Nov. 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2352484721001657

[6] Y. Zhang, D. Davis, and M. J. Brear, "The role of hydrogen in decarbonizing a coupled energy system," *Journal of Cleaner Production*, vol. 346, p. 131082, Apr. 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0959652622007144

[7] J. E. T. Bistline and G. J. Blanford, "Impact of carbon dioxide removal technologies on deep decarbonization of the electric power sector," *Nature Communications*, vol. 12, no. 1, p. 3732, dec 2021. [Online]. Available: http://www.nature.com/articles/s41467-021-23554-6

[8] C. Wilson, A. Grubler, K. S. Gallagher, and G. F. Nemet, "Marginalization of end-use technologies in energy innovation for climate protection," *Nature Climate Change*, vol. 2, no. 11, pp. 780–788, Nov. 2012. [Online]. Available: http://www.nature.com/articles/nclimate1576

[9] C. Wilson, A. Grubler, N. Bento, S. Healey, S. De Stercke, and C. Zimm, "Granular technologies to accelerate decarbonization," *Science*, vol. 368, no. 6486, pp. 36–39, Apr. 2020. [Online]. Available: https://www.science.org/doi/10.1126/science.aaz8060

[10] European Commission, "TIM analytics | Knowledge for policy." [Online]. Available: https://knowledge4policy.ec.europa.eu/text-mining/topic/tim_analytics_en

[11] Leiden University, "VOSviewer - Visualizing scientific landscapes," 2022. [Online]. Available: https://www.vosviewer.com//

[12] M. Reza, M. Mannan, S. B. Wali, M. Hannan, K. P. Jern, S. Rahman, K. Muttaqi, and T. M. I. Mahlia, "Energy storage integration towards achieving grid decarbonization: A bibliometric analysis and future directions," *Journal of Energy Storage*, vol. 41, p. 102855, Sep. 2021. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S2352152X21005788

[13] C. Camarasa, C. Nägeli, Y. Ostermeyer, M. Klippel, and S. Botzler, "Diffusion of energy efficiency technologies in European residential buildings: A bibliometric analysis," *Energy and Buildings*, vol. 202, p. 109339, Nov. 2019. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0378778819302129

[14] K. Jiang and P. Ashworth, "The development of Carbon Capture Utilization and Storage (CCUS) research in China: A bibliometric perspective," *Renewable and Sustainable Energy Reviews*, vol. 138, p. 110521, Mar. 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1364032120308066

[15] M. Ranjbari, Z. Shams Esfandabadi, F. Quatraro, H. Vatanparast, S. S. Lam, M. Aghbashlo, and M. Tabatabaei, "Biomass and organic waste potentials towards implementing circular bioeconomy platforms: A systematic bibliometric analysis," *Fuel*, vol. 318, p. 123585, Jun. 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0016236122004501

[16] A. Moro, E. Boelman, G. Joanny, and J. L. Garcia, "A bibliometric-based technique to identify emerging photovoltaic technologies in a comparative assessment with expert review," *Renewable Energy*, vol. 123, pp. 407–416, Aug. 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0960148118301599

[17] A. Moro, G. Joanny, and C. Moretti, "Emerging technologies in the renewable energy sector: A comparison of expert review with a text mining software," *Futures*, vol. 117, p. 102511, Mar. 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S001632872030001X