Proceedings of Global Power and Propulsion Society
ISSN-Nr: 2504-4400
GPPS Xi'an21
18th–20th October, 2021

www.gpps.global

GPPS-TC-2021-0286

DATA SCIENCE-BASED TREND ANALYSIS IN RESEARCH AND DEVELOPMENT OF TURBOMACHINERY

Mark Zieße
Jr. Research Group Multiphysics of Turbulent Flows
Institute of Turbomachinery and Fluid Dynamics
Leibniz Universität Hannover
Email: ziesse@tfd.uni-hannover.de
30823 Garbsen, Germany

Joerg R. Seume
Institute of Turbomachinery and Fluid Dynamics
Leibniz Universität Hannover
30823 Garbsen, Germany

ABSTRACT

Due to the very nature of science, the dissemination of knowledge is constantly increasing. This effect is reinforced by globalisation itself, as well as by the Internet. As shown in this publication, this is reflected in the disproportionate increase in the number of publications, making it more difficult for organisations to identify technologies and trends relevant to the sector, and to take them into account in their research and development agenda.
Both industry and research institutions are experiencing increased innovation pressure. This pressure is caused by globalisation-induced higher competition, as well as through requirements resulting from global challenges such as climate change. In addition, there are shorter product cycles and in particular an increase in computer-aided analysis methods.
In order to be able to compete internationally, it is necessary for both industry and research institutions to identify current market trends, new technical developments and current as well as future challenges, and to take these into account in their strategic orientation. Only then will it possible to advance the further development of turbomachinery in a targeted manner. The objective of this paper is to identify current market trends, technical developments and challenges in the context of turbomachinery; an application programming interface (API) was employed to analyse the database of the *www.elsevier.com* website.

INTRODUCTION

Although the development, design and global marketing of turbomachinery is a complex process and the barriers to entry are correspondingly high, the associated market is nevertheless subject to the creative destruction already recognized by Schumpeter (2008). According to (Schumpeter, 2008) innovation interrupts the usual economic cycle. As soon as a product is marketed and yields profits, there is an increased concentration of companies striving for market share. This increase in supply is followed by a reduction in prices and profit margins, which ultimately leads to the end of the product cycle once profitability is no longer given. Stopping this phenomenon and ensuring profitability is done through innovation. According to (Schumpeter, 2008) innovation can take five different forms: (1) introduction of a new product, (2) introduction of new production technologies, (3) development of new markets, (4) use of new technologies and materials, and (5) reorganization of production processes.
In order to be able to compete despite the creative destruction, companies in the aviation industry in particular rely on the process of open innovation (Richter and Walther, 2017). While the concept of open innovation was initially only observed in large multinational companies, this trend is also emerging in other companies (Chesbrough et al., 2003). The term "open innovation" was defined by Chesbrough et al. (2003) and describes the revealing of the innovation process of companies in order to increase their own innovation potential by strategically integrating the outside world. The source that creates knowledge does not necessarily have to correspond to the source of innovation and both sources do not have to originate from one company. Already in the year 1995 innovations received an input of knowledge from outside of the company by 34 to 65% (Conway, 1995). Open innovation offers 4 key benefits: (1) acceleration of value creation processes compared to competitors, (2) reduction of research and development costs, (3) higher market acceptance of the product due to a higher level of needs information from the customer, and (4) higher degree of innovation of the product (Piller et al., 2017). Due to

Table 1 Journal publication coverage by different data sources. The data sources are compared with regard to volumes of 6 different journals from the year 2009. The data were taken from Harzing (2019).

Reference journal	Data source					
	Crossref	Dimensions	Google Scholar	Microsoft Academic	Scopus	Web of Science
Academy of Management Journal	63	63	63	63	63	63
Journal of Accounting Research	40	40	40	40	40	40
Journal of Finance	78	78	78	67	78	78
Journal of International Business Studies	105	89	89	89	89	89
Journal of Marketing	61	61	61	61	61	61
Review of Economic Studies	69	50	50	68	50	50

increasing competitive pressure from globalisation, shorter product life cycles caused by digitalization and social challenges such as climate change, the innovation pressure is increasing, and more and more companies are opening up their innovation processes in order to be able to optimize their innovation processes. The opening of their innovation process also by not only purely large and multinational companies opens up opportunities for collaboration, especially for research institutions. The basic prerequisite of the open innovation process is therefore the formation of alliances or strategic partnerships between companies and research institutions (Allal-Chérif, 2015). For this, in turn, it is imperative to identify potential partners with complementary competencies. This is the starting point of the present work. The aim of the work is to identify companies and research institutions, their orientations and already existing alliances, as well as to recognize research trends. The latter is especially necessary for a target-oriented innovation management. For this purpose, all worldwide publicly available turbomachinery-specific publications were analyzed by applying data science methods.

METHODOLOGY

All results of this work and their visualisation were done by programmes written in the programming language Python in combination with open source extension packages. The stock market values shown are from the *https://finance.yahoo.com* website (yahoo!, 2021) and were accessed via the package *yfinance* (Aroussi, 2021).

Application Programming Interface

In the preliminary work for this paper, various APIs were tested. These include the API of *Springer* (Jones et al., 2021), the API of the non-profit organisation *Crossref* (Crossref, 2021), the API *Scopus* of the literature database of *Elsevier* (Elvesier, 2021) and the *Google Scholar* API (Google, 2021). Other databases or APIs to be mentioned are those of *Microsoft Academic* (Eide, 2020), *Web of Science* (Clarivate, 2021) and *Dimensions* (Dimension, 2021). After several tests, the *Scopus* API was chosen. The reason for this was that in *Scopus*, for example, authors and institutions are clearly defined and the different writing variations are directly assigned to them. Thus, in some cases (especially for the large institutions) up to more than 50 of different spelling variations were identified (including unconventional spellings). This feature of *Scopus* reduces the effort for data preparation enormously. Harzing (2019) have also shown that regardless of the database or API used, the volume of scientific publications is at a similar level. 6 different databases are listed with regard to their coverage of journal papers in Tab. 1. Regardless of the data source, the coverage of all data sources shown is comparable. The only exception is the API from *Springer*. This has shown a clearly low coverage regarding publications in own tests. Access to the *Scopus* API itself was achieved through the Python package *pybliometrics* (Rose and Kitchin, 2019).

Assignment of Papers to Turbomachinery

In all databases and APIs examined in this work, the papers and journals are subdivided according to the *All Science Journal Classification* (Elvesier, 2020). There is no further thematic classification of the papers. Therefore, a separate classification must be made with regard to turbomachines. For this purpose, a database query was first used to determine all papers that contain certain turbomachinery-specific terms either in the title or in the keywords. For these papers, the second step was to determine the journal or proceedings in which they were published. These journals or proceedings were then manually reviewed. In the case of journals or conferences that can be directly assigned to turbomachinery by their name, their publications were all assigned to turbomachinery and the database query was supplemented accordingly. Examples of such journals or conferences are e.g. the *Journal of Turbomachinery* or *Proceedings of Global Power and Propulsion Society*. The assignment is done both according to the specific journal or conference name, and to certain designations in the title and keywords in the text. If a publication was assigned to turbomachinery, its assignment to ASJC subareas was not taken into account. This limitation of publications according to turbomachinery is still necessary to reduce the number of data sets to be searched, as a search of the approx. 80 million publications is not practical in terms of time. The reasoning behind this decision is the limited query rate of search results from a query in the database. For the API used, for example, only a maximum of 20.000 database queries can be made in a week and 6 queries per second per API key (Wan, 2019).

Natural Language Processing

Natural Language Processing (NLP) describes the computer-aided processing of natural language in the field of computer science and artificial intelligence. In the context of this work, already implemented methods of the Natural Language Toolkit (NLTK) (Bird et al., 2009) were used. The titles of the publications were tokenized before further processing. Tokenization describes the decomposition of text sections into smaller linguistic building blocks. In this work, the titles were broken down into words. In a further step, capitalization and punctuations were removed and stop words were filtered out. Stop words are words that occur frequently but do not have a high relevance for capturing the text content. Examples are ”of” and ”by”.

Verification and Validation

For the verification within the scope of this work, two samples of 20 publications from the turbomachinery area and 20 papers from various areas were evaluated. For these papers, it was checked whether the associated return variables of the database used correspond to the intended variables and whether the returned values are correct. The comparison was always made with the original paper in pdf format. As expected, all return values of the 40 paper in total were correct and the access to the database via the API was thus verified, based on a reduced sample.

The validation, on the other hand, is more difficult. The validation should show that an arbitrary database query returns a number of results that is representative for all worldwide publications related to this query. The problem here is the availability of comparative data. No institution or database includes all worldwide publications. The idea is now to select a reduced number of data as comparative data for a validation. It is important that the reliability of these data is very high and that the number of publications is sufficiently large. Because of that, the publications of the authors' own institute were used as the comparative data basis for the validation. For this database, the authors can guarantee a high degree of reliability.

The starting point of the validation is a database query of all publications of which at least one author has *Institute of Turbomachinery and Fluid-Dynamics* listed as an affiliation, the affiliations country is *Germany* and the affiliation city is *Hannover*. Partial results of the validation are shown in Fig. 1. First, it was checked on the basis of the publication title whether the publications of the database query can be found in the reference data. A distinction was made between journal publications and proceedings publications. Of the total of 84 journal publications in the reference data, 71 are correctly recognised by the database, which corresponds to a relative share of 85%. For the proceedings publications, 151 of 213 are correctly recognised, which corresponding to a match of 71%. Secondly, it was checked whether the publication year

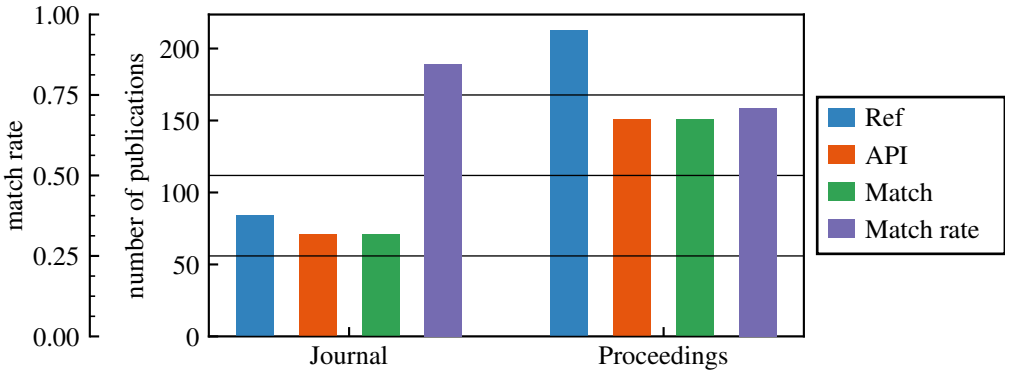


Figure 1 Results of the validation

and authors were correct between the reference and the database. No discrepancies were found here. In summary, it can be said that the database or API used returns the majority of all publications, and the metadata of the requested publications is correct. The agreement for the proceedings papers is significantly lower, but this is due to papers that are not available digitally. An example for such proceedings are the *Proceedings of International Gas Turbine Conference*.

RESULTS AND DISCUSSION

Overview

At the time of the query, the database contained 82 million records. Of these, 72% were Journal and 13% were proceedings papers. All results shown below consider only journal and proceedings papers. A breakdown of publications by subject is shown in Fig. 2. The overwhelming majority of publications can be assigned to the medical field. Approximately 12% comes from the engineering field. A further subdivision of this engineering field (see Fig. 2 on the right) shows that the major part of them can be assigned to mechanical engineering. The publications assigned to Turbomachinery according to the previous section account for only 2% of all publications. This share is at a similar level as for papers from the automotive (1%) and aerospace (3%) sectors. An illustration of the temporal development of the output of turbomachine-specific publications by country is given in Fig. 3. There is a shift in publication output from North America and Europe to Asia.

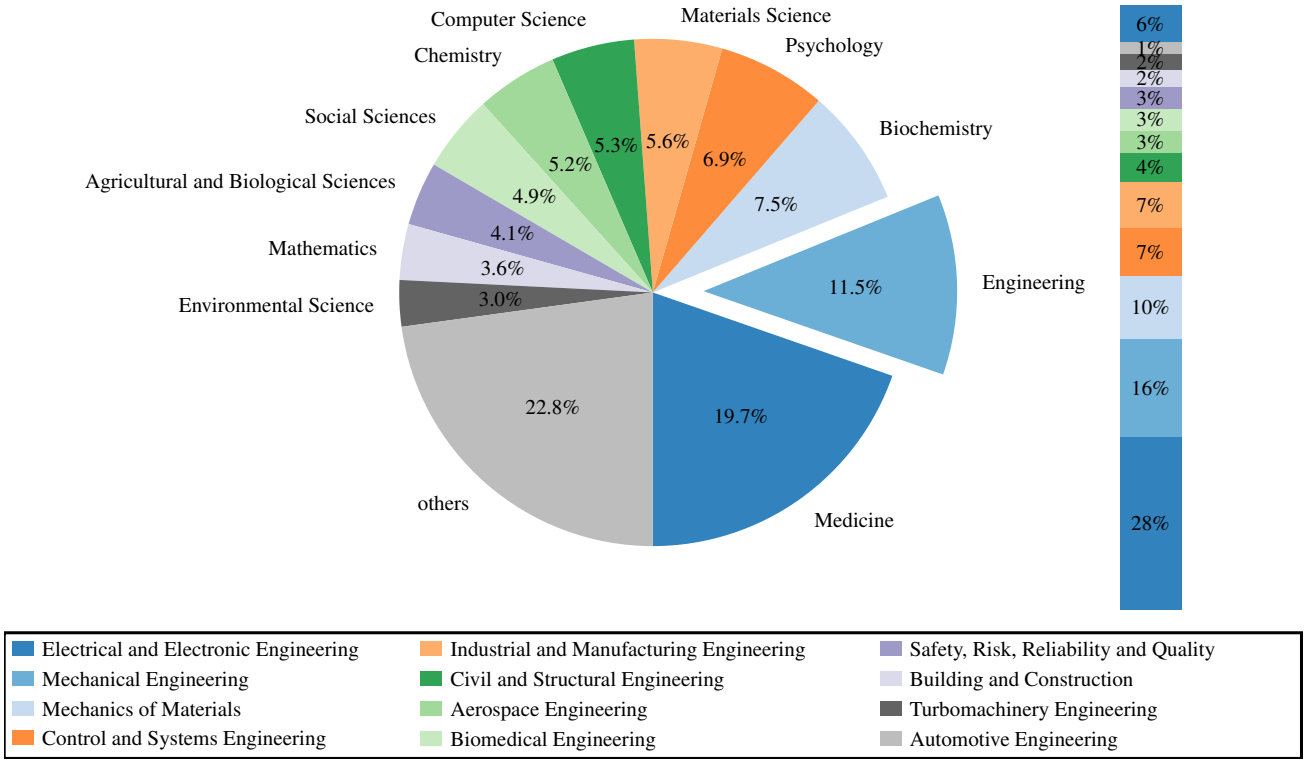


Figure 2 Subject classification of all journal and proceedings publications in the database (left) and further subdivision of publications from the engineering sector (right).

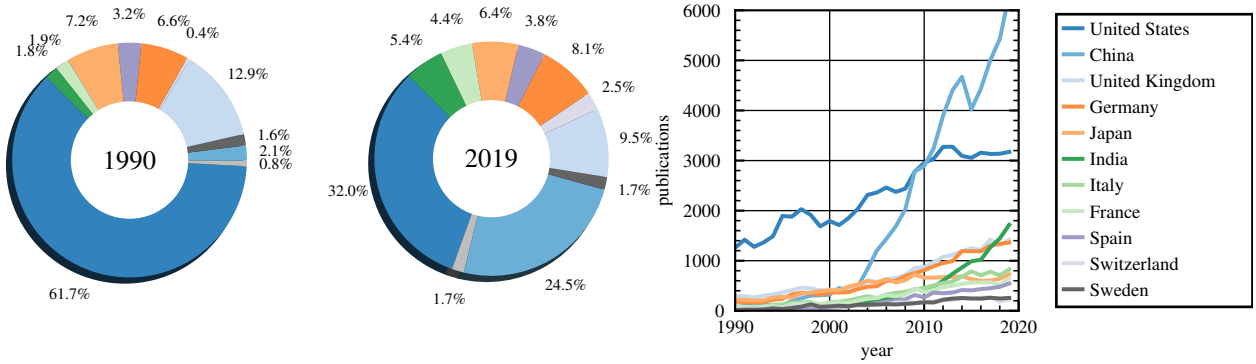


Figure 3 Temporal development of the annual publication output of the 10 countries with the largest number of publications (right) and comparison of the distribution of all publications up to 1990 and 2019 by these 10 countries (pie charts).

While the European countries and the USA show a constant linear growth in annual publications, a disproportionate increase since 2003 can be observed for the annual publication output of China. India has also seen disproportionate growth since around 2012, although this is not as significant as in China. On the other hand, the volume of publications has stagnated for Japan since 2010. The disproportionate growth of China in particular has also a clear influence on the relative distribution of all turbomachinery-specific publications by country. While China's share of all turbomachinery-related publications was still around 2% in 1990, it is already 24.5% in 2019. A similar increase in the relative share can be seen for India. For India, the rel. share increases from 1.8% to 5.4%. If the increase in publication output from China continues, in a few years the largest share of annual turbomachinery-specific publications and also the largest share of all existing publications will be attributable from China.

Perspective Research Institutes

For the total of approx. 400.000 publications in the field of turbomachinery, a total of over 70.000 different participating institutions were identified. Below, only institutions which were involved in at least 5 publications were considered. This

reduced the number of institutions to 9908. The share of public institutions was 59% and of companies 41%. Figure 4 top shows the development over time of the annual publications of the 10 companies with the largest number of publications in the context of general turbomachinery. In addition, the relative share of these companies in all publications published to date in 1990 and 2020 is shown. *General Electric (GE)* by a significant margin annually publishes the most, maintaining a slight yearly increase. Except for a particularly good 1997 for Rolls-Royce, the annual publication output of *Siemens* and *Rolls-Royce* is comparable. Since about 2010, however, the annual publications for both have not increased as much as in the years before; for the last few years a trend towards stagnation is observable. For *Alstom*, a complete collapse of publications after 2015 can be observed. This can be explained by the takeover of the *Alstom* energy division by *GE* (GE, 2015). For the other companies shown, however, the publication output is at a rather constant level or even slightly decreasing for the last decade. Regardless of the firm considered, an increase would actually be expected, as observed for the various countries (see Fig. 3). One reason for this is the diversification of publication output across more firms, which is also confirmed by the distribution of all publications in 1990 and 2019 (Fig. 4 below). While *GE* accounted for over 10% of all publications in 1990, this share is slightly above 5% in 2019. Additionally, a significant change in the top 10 companies with the most publications can be seen between 1990 and 2019, with only 4 of the 10 companies shown from 1990 still represented in 2019. Another reason for the stagnation of the annual publications of the companies shown could be that other participating institutions publish results on their own, according to the open innovation process or the companies may invest fewer resources in the direct research and development of turbomachinery. Figure 5 shows a comparison of the

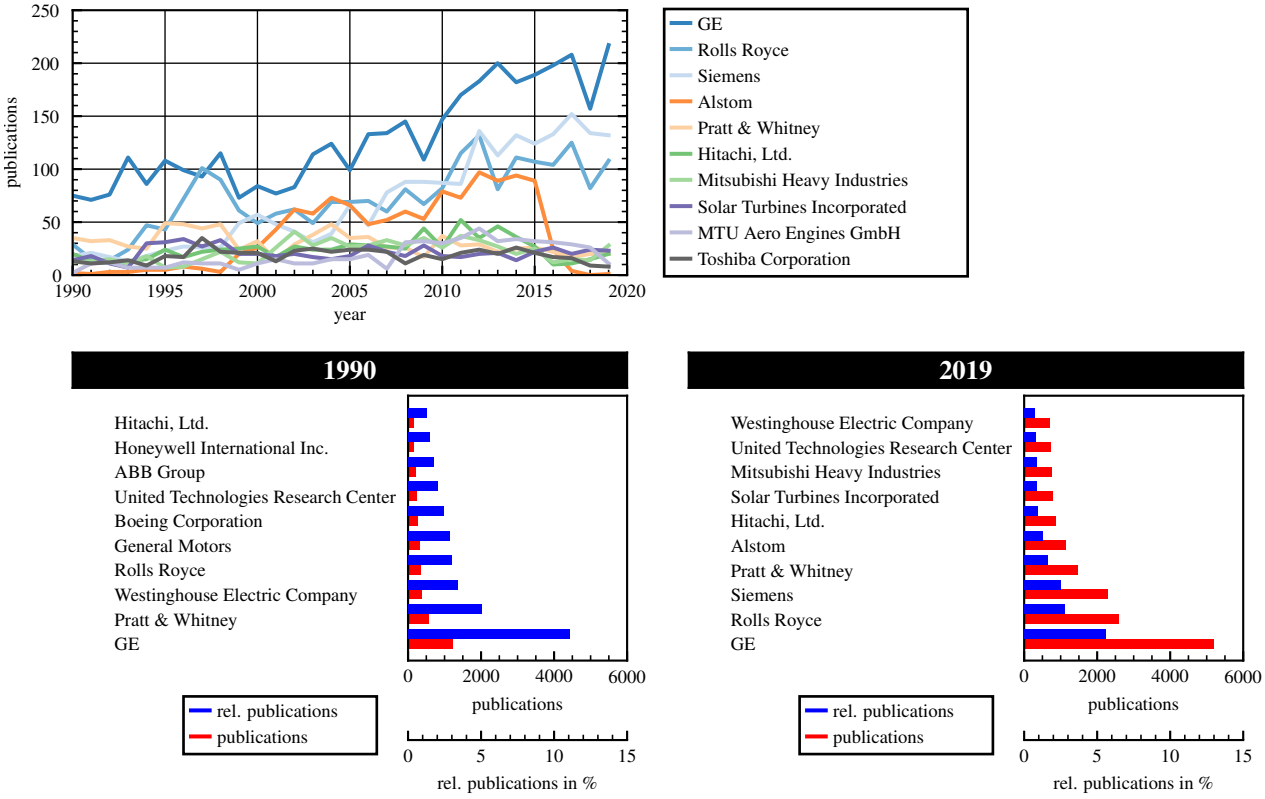


Figure 4 Time evolution of the annual publication output of the current 10 **companies with the most publications (top) and the distribution of all publications among these** 10 **companies in** 1990 **(bottom left) and** 2019 **(bottom right).**

publication rate with the stock market value of the companies from Fig. 4. Also shown is the normalized cross-correlation of publication rate and stock market value. If the cross-correlation would correspond to a step function with the value 1 for a phase lag of zero, the relative course of stock market value and publications would be identical. The idea behind the consideration of the cross-correlation is to analyze whether there is a direct correlation for a quantity describing the financial situation of a company, such as the stock market value, and the output of publications. A positive phase lag means that the stock market value is lagging behind the publications and a negative value means the opposite. Looking at the cross-correlation does not show a clear trend across all companies. For most companies, the cross-correlation approximates a triangular function with a maximum value at around zero. This suggests a correlation between stock market value and publication rate, but without a phase offset. Fig. 6 (center left) exemplifies the relative share of publications without the collaboration of a partner of 5 companies with the highest number of publications. This rate is highest for *GE* with almost 30%. *Siemens* has a comparable own publication rate of around 30%; *Rolls-Royce* and *Pratt & Whitney* stand at around 10%. Thus, the proportion of publications without the participation of a partner is about 10% to 40%, which conversely

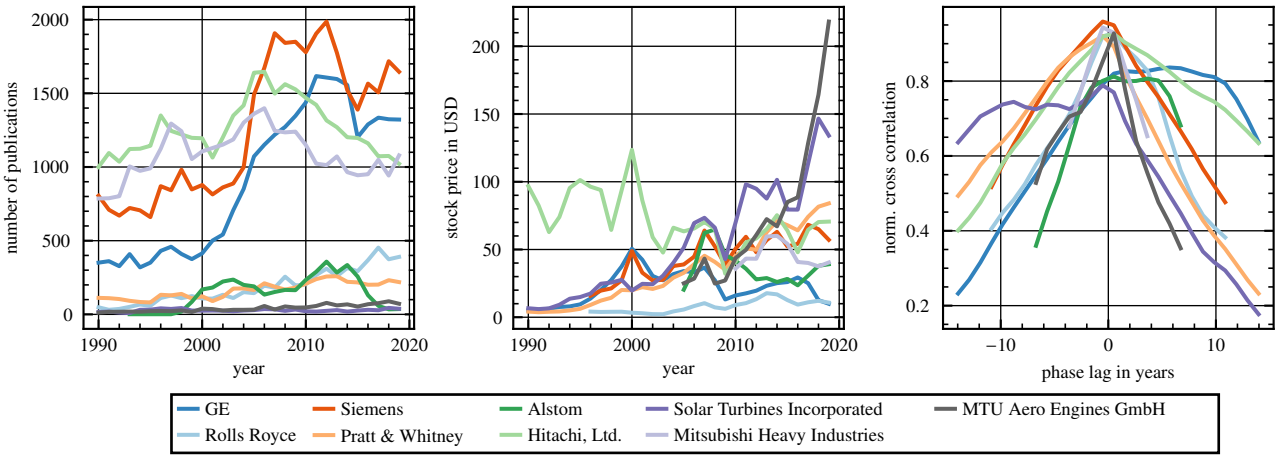


Figure 5 Comparison of the publication rate (left) with the stock market value (center) of the companies with the highest publication rates in 2019 **and normalized cross-correlation of publication rate and stock market value (right).**

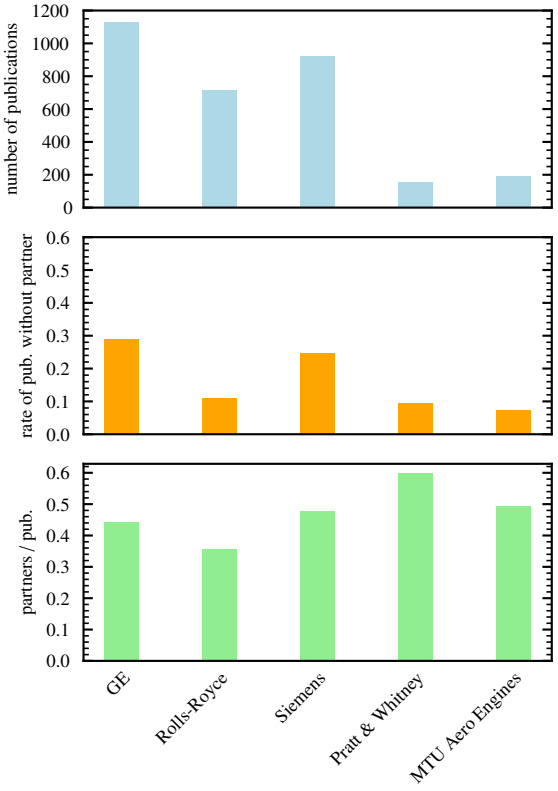
means that over 60% of all publications are co-written with the participation of another institution. The number of partners to the number of publications (fig. 6 bottom left) varies between about 35% and 60%. *Pratt & Whitney* and *MTU Aero Engines* have the most partners relative to publications at about 50 – 60% while have a low number of publications with no partners involved. Relatively speaking, both companies collaborate most extensively with other institutions. With regard to the most frequent partners in publications (see fig. 6 on the right), clear trends can be identified. GE collaborates by far the most with the *Universitâ degli Studi di Firenze*. *Rolls-Royce* collaborates with U.K. universities. *Siemens*, on the other hand, shows a high degree of diversification in terms of partners for publications. For *Pratt & Whitney*, the most frequent research partner is the *Pennsylvania State University*. *MTU Aero Engines* has most frequently collaborated with the *Aristotle University of Thessaloniki* and the *Deutsches Zentrum für Luft- Und Raumfahrt*. In summary, the number of publications without participation of further institutions is still very high, up to 40%, and certain patterns (of collaboration) are noticeable between companies and partners.

Perspective Companies

As Fig. 7 shows, the top 5 of the top 10 public institutions in terms of publication rate are from China. For all these universities, the publication rate has a similar level and a steady increase can be seen, which agrees well with the results in Fig. 3. Here, the *Tsinghua University* has the highest publication rate. Together with *Harbin Institute of Technology* and *North China Electric Power University*, 7 of the top 10 are from China. It is noticeable that the publication rate of all Chinese institutions is increasing, while that of the other institutions remains constant or even decreases. For example, *Danmarks Tekniske Universitet* shows a stagnation of publications per year after the year 2014, and *NASA Glenn Research Center*'s annual publications have been decreasing slightly since the year 1996. The publication rate of the *Jet Propulsion Laboratory* has stagnated since about 2005. A comparison of the distribution of publications in 1990 and 2019 shows, on the one hand, a diversification, and on the other hand, a clear shift from North America and Europe to China. For example, in 1990, the *NASA Glenn Research Center* accounted for the largest share of publications by a significant margin at just under 4%. In 2019, Tsinghua University accounts for the largest share, at about 1%. Comparing the top publishing companies (see Fig. 4) with the top publishing public institutions (see Fig. 7), it can be seen that Chinese public institutions account for the largest share of new publications, but there is no Chinese company in the top 10. This means that, currently, a lot of knowledge is being generated in China, but it is not finding its way into national companies. This offers foreign companies the opportunity to include unaffiliated Chinese public research institutions in their research.

Trend Analysis

Among the associated titles and keywords of all papers assigned to turbomachinery, the most frequently recurring words and word-pairs were determined, in order to identify trends by their relative annual increase. The results of this are shown in Fig. 8 and compared with the annual relative number of publications from the turbomachinery sector. There is a clear increase in publications with neural networks already since 1990. This increase is even more pronounced since 2005, and in the last two to three years an exponential growth is seen. The use of neural networks is thus a methodological trend in the field of turbomachinery. Another technological trend seems to be the exploration of hydrogen. Here, a significant increase can be seen around the year 2000. Since 2009, composites also seem to be moving closer into the focus of research, as well as questions of maintenance. In addition, also since about 2000, there has been an above-average and similar



GE	
Università degli Studi di Firenze	9.2%
Rheinisch-Westfälische Technische Hochschule Aachen	2.5%
Politecnico di Milano	2.5%
Technical University of Munich	2.4%
University of Melbourne	1.8%
others	81.5%

Rolls-Royce	
University of Oxford	9.7%
University of Cambridge	7.7%
Imperial College London	6.6%
Brandenburgische Technische Universität Cottbus	4.4%
Cranfield University	3.2%
others	68.5%

Siemens	
Linköpings universitet	3.7%
Danmarks Tekniske Universitet	3.0%
Universität Stuttgart	2.8%
Universität Duisburg-Essen	2.4%
Lunds Universitet	2.0%
others	86.1%

Pratt & Whitney	
Pennsylvania State University	20.9%
United Technologies Research Center	6.2%
Polytechnique Montréal	6.2%
École de Technologie Supérieure	4.9%
Politechnika Rzeszowska	3.6%
others	58.2%

MTU Aero Engines	
Aristotle University of Thessaloniki	8.7%
Deutsches Zentrum fur Luft- Und Raumfahrt	6.7%
Universität Stuttgart	6.1%
Rheinisch-Westfälische Technische Hochschule Aachen	5.4%
Universität der Bundeswehr München	4.5%
others	68.6%

Figure 6 Number of publications (top left), relative rate of publications without partners (middle left) and ratio of number of partners to publications (bottom left) for the last 5 **years of** 5 **companies with the highest number of publications. Additionally shown are the** 5 **most frequent partners of the respective company in publications (right).**

increase in publications dealing with stability and non-linearity. For the word ”vibrations”, a comparable trend can be seen (but not shown here). This indicates that aeroelastic issues with a focus on stability (such as forced response or flutter) are also increasingly being investigated. In addition, there is a significant increase in publications on acoustic topics, and monitoring seems to have become more important since about 2005, as well as investigations on cooling. When comparing the publications with an experimental and numerical focus, it is noticeable that, as expected, the numerical investigations have increased more significantly than the experimental ones, but an above-average increase can also be observed for the experimental investigations. The numerical investigations are thus not carried out instead of, but in addition to, the experimental ones. The trend analysis was also carried out according to the type of turbomachine or component. A distinction was made between gas and steam turbines, as well as turbines in aircraft engines (hereinafter referred to as ”turbine”), wind turbines, axial compressors and axial blowers (hereinafter referred to as ”compressor”), any type of radial engine including turbocharger (hereinafter referred to as ”turbocharger”), and fans.

Figure 9 on the left shows the annual absolute number of publication by component or type. Turbocharger publications account for the smallest share and wind turbine publications for the largest. Turbine and compressor are at a similar level, with the annual number of publications of the turbine being generally larger. There is also a clear increase in wind turbine publications since 2000, and this increase is even more evident when looking at the relative publication rate (see Fig. 9 middle). There is a significant increase in wind turbine publications that exceeds the general turbomachinery publication rate by a factor of about 6. However, this increase appears to be flattening out. For turbochargers, there had also been an increase compared to the general turbomachinery trend since about 2005, but a stagnation can be observed here in recent years. For compressors and turbines, there is an increase in annual publications, but this is lower than the trend for all publications. The rel. increase in publications for fans lies between turbines and compressors. The distribution of all publications by component and type over time (see Fig. 9 right) shows an increase of the wind turbine share from about 7% to 19% at the cost of the turbine and partly the compressor. The turbocharger share increases over time, but remains very small. Up until the year 2010, the share of fans had been slightly increased up to around 12%. This (relative) share has now kept constant. Fig. 10 shows for each component or type the relative development over time of the count of the 4 most frequent word pairs in titles and keywords. Additionally shown are the most frequent word pairs with numerical and

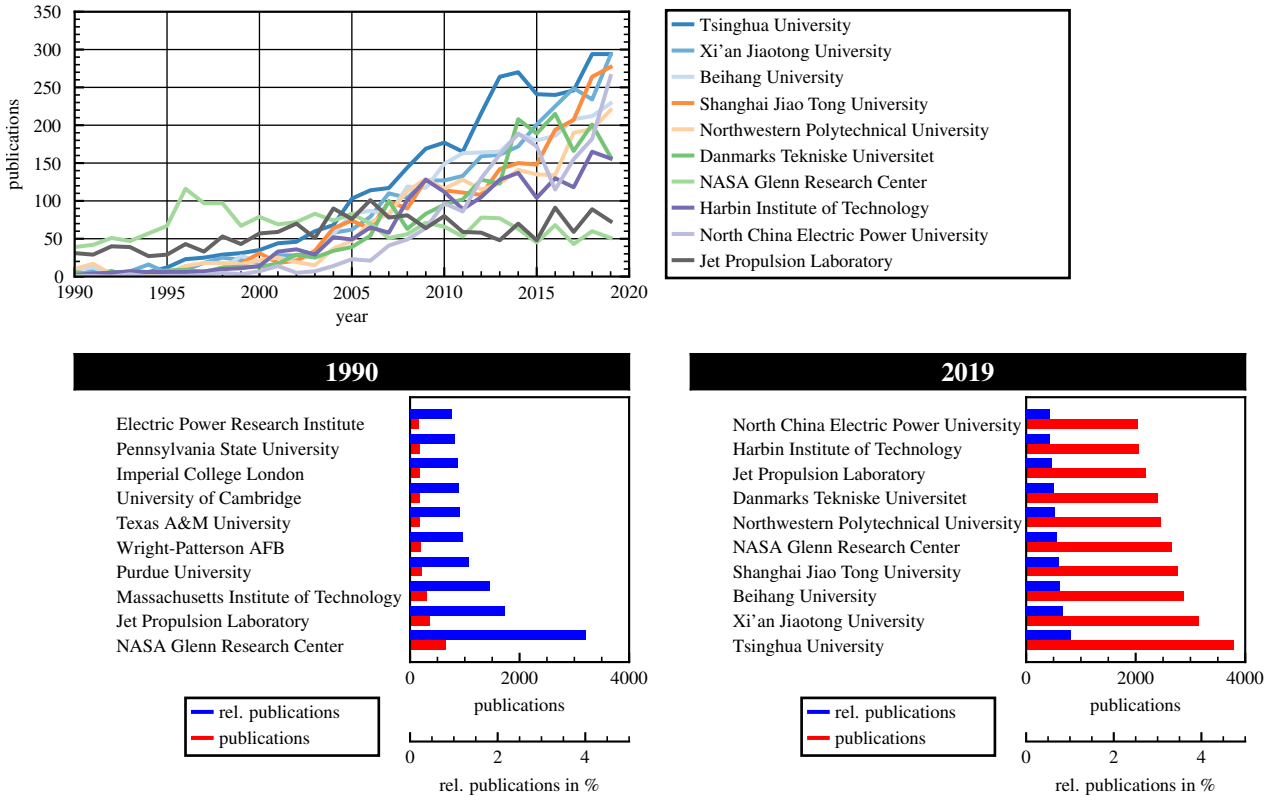


Figure 7 Time trends in annual publication output of the current top 10 **publishing public institutions (top) and the distribution of all publications among the top** 10 **publishing public institutions in** 1990 **(bottom left) and** 2019 **(bottom right).**

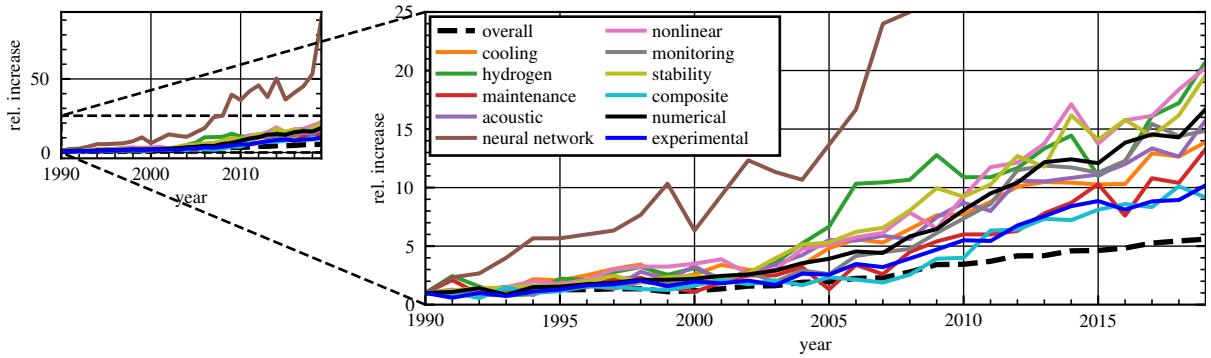


Figure 8 Comparison of the rel. increase of all turbomachinery publications with rel. increase of publications with specific words or word pairs in the title or the keywords that occurred most frequently in the last 5 years. In addition, the increase of publications with numerical or experimental background is shown. The values shown are relative to 1990**.**

experimental background, as well as the relative publication rate of the considered component or type. For wind turbines, there is a significant increase in publications in the offshore sector, as well as an above-average increase in publications with floating wind turbines. For turbines, two significant trends can be identified. On the one hand, publications with numerical focus increase above average, on the other hand, micro turbines seem to become more in focus. For the compressor, casing topics in particular seem to have gained above-average importance since 2000. For the turbocharger, all considered word pairs are below the general relative publication rate, except for the word pair "Diesel engine". For this, an increase could be seen until about 2014, followed by a significant decrease to a level below the general relative publication rate. According to the authors, this decrease is a direct result of Dieselgate. For fans, there is an above-average increase for all word pairs shown. In particular, noise reduction has gained strongly in importance since about 2002, as well as investigations of fans regarding broadband noise.

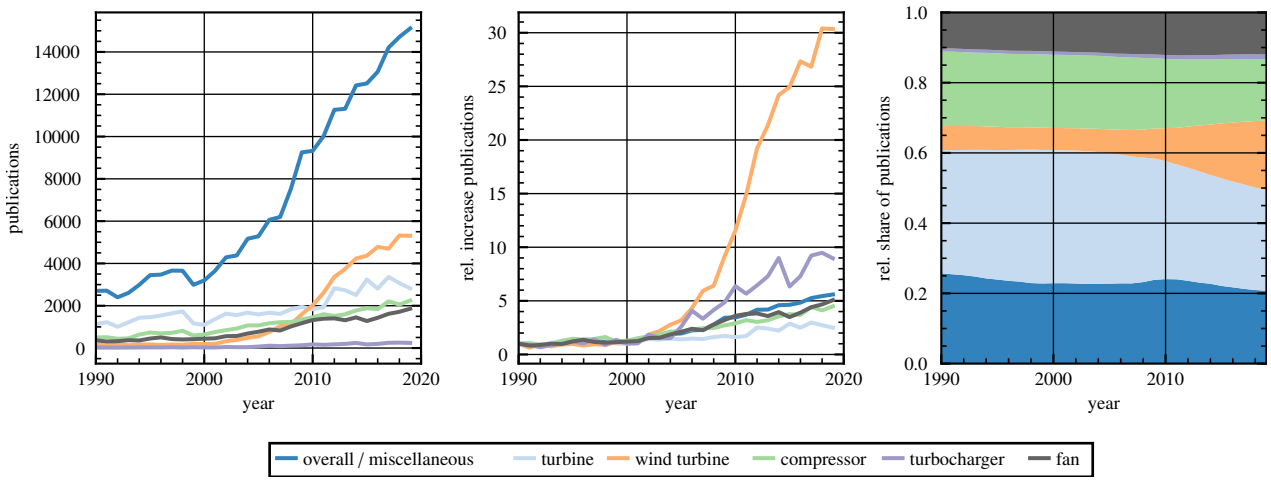


Figure 9 Comparison of all turbomachinery publications with publications by component or type over time (left), relative annual increase in publications normalized to 1990 **(center), and relative distribution of all publications by component or type over time (right).**

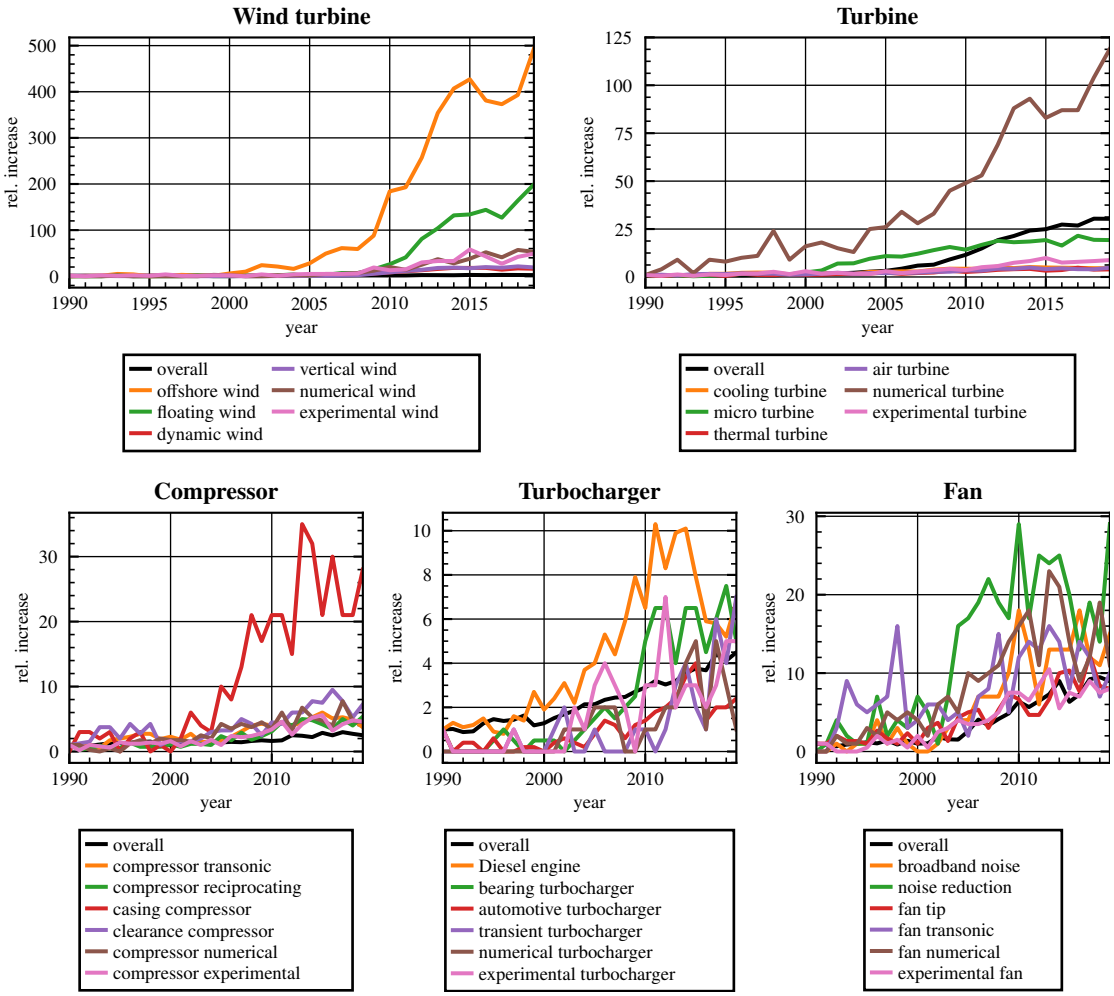


Figure 10 Relative time evolution of the count of the 4 **most frequent word pairs in titles and keywords for each component or type. Additionally, the most frequent word pairs with experimental and numerical background are shown, as well as the relative increase of all publications of the respective component or type. The normalization was done by the year** 1990**.**

CONCLUSIONS

Various methods of data science were applied to the worldwide publications on turbomachinery and wind turbines, showing ...

- Of all the publications worldwide, 12% can be assigned to the engineering field. Of these, only about 2% have a turbomachinery background.
- The annual output of publications is by far the largest for China and continues to increase. In a few years, the largest share of all existing turbomachinery-specific publications will probably be attributable to China.
- The companies with the highest publication output are currently still from North America and Europe, with the annual publication rate of the companies stagnating. The stock prices of these companies correlate with the publication rate without significant time lag.
- In the case of publications by public institutions, a shift from North America and Europe to China and, in addition, a diversification can be seen.
- Through a trend analysis, clear methodical and technical trends as well as investigated challenges could be identified. For example, there is a continuous above-average increase in publications focusing on hydrogen, aeroelasticity, and acoustics, as well as publications on maintenance, cooling, composites, and monitoring. In methodology, there is an exponential increase in the use of neural networks, and an above-average increase in publications with an experimental background.
- There is an above-average increase in publications in wind turbines, as well as an exponential increase in publications on turbomachinery with a numerical background for the turbine. Clear trends on different turbomachinery components and types emerge.

Now that the general process chain has been applied for this publication, further data, e.g. impact factors or patents, will be used for identifying further correlations.

REFERENCES

Allal-Chérif, O. (2015), 'The way towards open innovation: Airbus multi-functional teams', *European Scientific Journal, ESJ* **11**(10).

Aroussi, R. (2021), 'yfinance package'. [Online; accessed 6th April 2021].
URL: *https://pypi.org/project/yfinance/*

Bird, S., Klein, E. and Loper, E. (2009), *Natural language processing with Python: analyzing text with the natural language toolkit*, " O'Reilly Media, Inc.".

Chesbrough, H., Press, H. B. S. and Brown, J. (2003), *Open Innovation: The New Imperative for Creating and Profiting from Technology*, Harvard Business School Press.

Clarivate (2021), 'Web of science api'. [Online; accessed 6th April 2021].
URL: *https://developer.clarivate.com/apis*

Conway, S. (1995), 'Informal boundary-spanning communication in the innovation process: an empirical study', *Technology Analysis & Strategic Management* **7**(3), 327–342.
URL: *https://doi.org/10.1080/09537329508524216*

Crossref (2021), 'Crossref api'. [Online; accessed 6th April 2021].
URL: *https://github.com/CrossRef/rest-api-doc*

Dimension (2021), 'Dimensions api'. [Online; accessed 6th April 2021].
URL: *https://www.dimensions.ai/dimensions-apis/*

Eide, D. (2020), 'Microsoft academic api'. [Online; accessed 20th April 2021].
URL: *https://docs.microsoft.com/en-us/academic-services/project-academic-knowledge/introduction*

Elvesier (2020), 'All science journal classification'. [Online; accessed 6th April 2021].
URL: *https://service.elsevier.com/app/answers/detail/a_id/15181/supporthub/scopus/*

Elvesier (2021), 'Scopus api'. [Online; accessed 6th April 2021].
URL: *https://dev.elsevier.com/sc_apis.html*

GE (2015), 'Ge completes acquisition of alstom power and grid businesses'. [Online; accessed 19th April 2021].
URL: *https://www.ge.com/news/press-releases/ge-completes-acquisition-alstom-power-and-grid-businesses*

Google (2021), ‘Google scholar api’. [Online; accessed 6th April 2021].
URL: *https://serpapi.com/google-scholar-api*

Harzing, A.-W. (2019), ‘Two new kids on the block: How do crossref and dimensions compare with google scholar, microsoft academic, scopus and the web of science?, scientometrics’, *Scientometrics* **120**(1), 341–349.

Jones, E., Oliphant, T., Peterson, P. et al. (2021), ‘Springer nature api portal’. [Online; accessed 6th April 2021].
URL: *https://dev.springernature.com/*

Piller, F. T., Reichwald, R., Möslein, K. and Ihle, C. C. (2017), *Interaktive Wertschöpfung kompakt: Open Innovation, Individualisierung und neue Formen der Arbeitsteilung, Lehrbuch*, Springer Gabler, Wiesbaden.

Richter, K. and Walther, J. (2017), *Supply chain integration challenges in commercial aerospace: a comprehensive perspective on the aviation value chain*, Springer, Cham.

Rose, M. E. and Kitchin, J. R. (2019), ‘pybliometrics: Scriptable bibliometrics using a python interface to scopus’, *SoftwareX* **10**, 100263.
URL: *https://doi.org/10.1016/j.softx.2019.100263*

Schumpeter, J. A. (2008), *Capitalism, Socialism, and Democracy - Third Edition*, HarperCollins, Heidelberg.

Wan, K. (2019), ‘Introduction to scopus apis’. [Online; accessed 6th April 2021].
URL: *https://www.elsevier.com/de-de/events/scopus-community-anwendertreffen?a=917179*

yahoo! (2021), ‘yahoo! finance website’. [Online; accessed 6th April 2021].
URL: *https://finance.yahoo.com*