

Theory and Practice on Non-Probabilistic Data and Analysis: a bibliometric review

Jeanfrank Teodoro Dantas Sartori*

Abstract. This bibliometric study aims to summarize the academic landscape of non-probabilistic data research, based on an examination of scientific output indexed in Web of Science and Scopus databases. It employs multiple methods to analyse and describe the collected corpus, including co-authorship and keyword co-occurrence networks to investigate patterns of collaboration and predominant research themes. Co-authorship analysis identified several robust research clusters, while keyword later spotlighted key thematic areas in the field. Countries, types of documents, categories, year of publication, citations and other metrics were also produced, and implications discussed. The findings present a structured overview of the non-probabilistic data research landscape, delineating the research trends, prominent authors, and emerging themes.

Keywords: Non-Probabilistic, Non-Random, Statistics, Bibliometric, Review

1. Introduction

Over the past several decades, a variety of methods to process and analyze data have been developed and applied, constitution an essential role in numerous fields. Non-probabilistic data analysis is one such approaches that, despite its relative novelty, has found substantial application in various domains, operating independently of probabilistic assumptions, which are commonly used in traditional data analysis and statistics [42]. It embraces the vagueness and imprecision inherent in many real-world scenarios, thereby making it an effective tool in modelling and decision-making processes [11].

As a key associated technique, non-random sampling, contrary to the probabilistic analog, incorporates elements like convenience, judgment, or purposiveness in the selection process [18] [37] [7]. Despite criticisms concerning its representativeness and inference power, it has found considerable application, especially in exploratory research, case studies, and situations where random sampling may not be feasible or cost-effective [4], with recent approaches addressing confidence level and margin of error metrics for those scenarios [31].

The inherent subjectivity in non-random sampling adds to the uncertainty in non-probabilistic data analysis, forming a dimension of flexibility and adaptability to research

* Business School, Positivo University, Curitiba, Brazil (contatto@jeanfrank.it)

design [1]. Nevertheless, has found extensive use in uncertainty quantification and risk management [5]. Through methodologies like fuzzy logic, possibility theory, and evidence theory, non-probabilistic data analysis provides nuanced perspectives on risk and uncertainty that extend beyond traditional probabilistic frameworks [11] [32].

Its versatility of is also reflected in its applications spanning fields such as artificial intelligence, engineering, environmental science, and health sciences, among others [44] [16]. In these contexts, the tolerance for imprecision and uncertainty has enabled it to generate insightful and robust models for complex, real-world scenarios [2].

In recent years, the rise of big data and the advent of more sophisticated computational tools have further accentuated the relevance of non-probabilistic data analysis [40]. As the volume and complexity of data continue to escalate, it can offer valuable strategies to interpret and model this information in a manner that acknowledges its inherent uncertainties and ambiguities [20].

In the field of machine learning, particularly through fuzzy logic, has been instrumental in developing algorithms that can handle imprecise and vague data [43]. Fuzzy systems, with their capacity to handle linguistic variables and subjective human reasoning, have been applied extensively in pattern recognition, control systems, and prediction models [17].

Despite the growing recognition of non-probabilistic data analysis, comprehensive reviews and bibliometric analyses on the subject are scarce [46]. A bibliometric review can provide valuable insights into the development, trends, and structure of a research field, offering a holistic picture of its progression over time [12].

Therefore, this study has the following research question: how the current state-of-the-art of scientific knowledge on non-probabilistic data and analysis can be described? To answer it, a bibliometric review was conducted on the subject by examining key publications, authors, and thematic trends, mapping the evolution of both theory and applications, thereby contributing to our understanding of this growing significant research area and its ability to interpret and model the increasingly complex data landscapes of the 21st century [19].

2. Methodology

This study conducted a comprehensive bibliometric review, a well-established method for quantitatively analyzing scientific literature in a given field [12]. It not only shed light on the intellectual structure of a research field but also map its evolution, identify key contributors, and uncover prevalent themes and trends [48]. By employing such approach, it is intended to map the landscape of non-probabilistic data analysis and its applications and understand its academic contributions over time.

For data collection, this study considered Web of Science and Scopus databases, which are among the largest and most reliable databases for scientific literature [26]. The search was conducted using the keywords "Non-Probabilistic," "Non Probabilistic," "Non Random Sample," and "Non-Random Sample," encompassing a broad spectrum of research related to non-probabilistic data analysis, as shown on table 1. Those search strings were applied to the title, abstract, and keywords of articles, ensuring that the returned publications were of direct relevance to the topic.

No filter was applied in terms of year of publication, since the research intended to show the broad perspective of the subjective also chronologically. A total of 3,458 publications

was retrieved from Web of Science, while 392 publications from Scopus, resulting in a total of 3,850, of which 3,340 after duplications filtering.

Table 1. Search criteria

Web of Science	(TS="Non Probabilistic") OR (TS="Non-Probabilistic") OR (KP="Non Probabilistic") OR (KP="Non-Probabilistic") OR (TS="Non Random Sample") OR (TS="Non-Random Sample") OR (KP="Non Random Sample") OR (KP="Non-Random Sample")
Scopus	TITLE("Non Probabilistic") OR KEY("Non Probabilistic") or TITLE("Non-Probabilistic") OR KEY("Non-Probabilistic") or TITLE("Non Random Sample") OR KEY("Non Random Sample") or TITLE("Non-Random Sample") OR KEY("Non-Random Sample")

Primarily, co-authorship networks were scrutinized to investigate the collaborative patterns among researchers, the strength of co-authorship and the contemporaneity of researchers. Alongside, keyword co-occurrence networks were analysed to visualize the thematic structure of the field, uncover prevailing research trends, and pinpoint emerging topics of interest [9]. The strength of their co-occurrence provided insights into the prominent themes within non-probabilistic data analysis research.

The collected data and ensuing networks were visualized and analysed using VOSviewer, a software tool specifically designed for constructing and viewing bibliometric maps [38]. Its capacity to depict large bibliometric networks, combined with its clustering algorithms, made it an adequate tool for portraying the structure of them.

3. Results and discussion

Breaking down of the publication types in the collected corpus, as shown on figure 1, reveals that conference papers constitute the largest category, accounting for 51.4% of the total, what might suggest a rapidly developing state, as they are typically quicker to produce and publish than journal articles, allowing researchers to disseminate their findings promptly, fostering quicker feedback and collaboration. They also indicate the presence of numerous conferences related to the topic – or applying it.

Journal articles, which make up 41.8% of the corpus, are the second most common publication type, that requires rigorous peer-review processes and often hold high prestige in academia indicating a robust foundation of the theme. It suggests a maturing field, where research has been extensively scrutinized and has reached levels of refinement and consolidation to merit publication in scholarly journals. It also signals that these themes have attracted considerable attention from academia and, likely, funding sources. This correlates with the finding of the citation analysis discussed further on, as the majority of top-cited papers were of this type category.

The remaining publication types comprise a smaller fraction of the corpus, with early access publications accounting for 2.6%, while reviews indicate that sufficient work has been conducted to merit synthesis and assessment of the current state of knowledge, suggesting a continually expanding and deepening understanding of the issues at hand. Other types like book chapters and editorials, though scarce, represent an effort to synthesize, communicate,

and discuss ideas in a more comprehensive or conversational format, reaching beyond traditional research articles.

The corpus also includes a small fraction of less common types such as letters, retracted papers, conference reviews, notes, short surveys, data papers, and meeting abstracts. The diversity of these types, though individually representing a small portion of the corpus, contributes to the richness and multiplicity of academic communication forms. As the corpus continues to grow, it would be insightful to track the evolution and distribution of these publication types to reveal trends and shifts in the scholarly discourse around these themes.

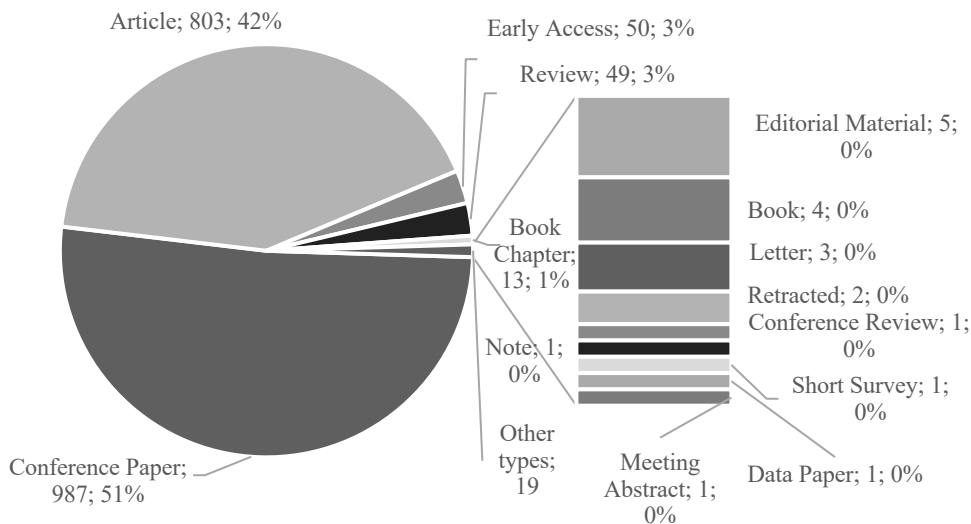


Figure 1. Number of Publications per Type

From the analyzing according to their subjects, as shown on table 2, it must be highlighted that the standards differ a bit between the databases, however the broad trends in the data still provide useful insights into the landscape of research on the themes. The highest proportion falls under Engineering (9.8%), closely followed by Computer Science (7.0%), and Mathematics (5.6%). This finding resonates with the countries with the highest number of publications, as in-depth discussed further on, where countries known for their strong emphasis on STEM disciplines (China, United States) led the list. The prominence of these categories also indicates that the research themes are deeply rooted in technical and quantitative disciplines, suggesting that computational, mathematical, and engineering techniques play a significant role in the research.

Education/Educational Research (3.3%) and Public Environmental Occupational Health (2.6%) are among the other prominent categories. This might point to a growing intersection of the technical disciplines with educational and health-related applications and could suggest that the research is not just technical but also has practical implications for education and health sectors.

Other categories include Mechanics, Multidisciplinary Engineering, Computer Science Artificial Intelligence, Mechanical Engineering, and Multidisciplinary Psychology, among others, that further highlight the inter- and multidisciplinary nature of the research themes. It

is also worth noting the presence of categories like Nursing, Environmental Sciences, and General Internal Medicine. Future studies might further investigate these interdisciplinary patterns and their implications for the development of the field.

Table 2. Top 20 Categories/Subjects in Number of Publications

Category/Subject	Publications
Engineering	750
Computer Science	536
Mathematics	431
Education Educational Research	248
Mechanics	205
Public Environmental Occupational Health	199
Engineering Multidisciplinary	193
Computer Science Artificial Intelligence	179
Engineering Mechanical	163
Psychology Multidisciplinary	163
Engineering Electrical Electronic	159
Physics and Astronomy	145
Computer Science Theory Methods	144
Computer Science Interdisciplinary Applications	115
Mathematics Interdisciplinary Applications	114
Management	109
Computer Science Information Systems	105
Nursing	104
Environmental Sciences	104
Medicine General Internal	102

In the analysis per countries, as shown on figure 2 and table 3, China leads with 1051 articles, what is consistent with its increasing prominence in global scientific research and its recognized commitment to research and development, evidenced by its heavy investment in this sector. This dominance can also be linked to the frequent occurrence of Chinese authors, such as Wang and Liu, as analyzed later. The United States, with 529 articles, comes next, reflecting its strong tradition of scientific research and its well-funded research institutions, universities, and industries. Spain, Brazil, and Mexico, each with significant numbers of articles, follow. These countries, along with Portugal, Chile, Colombia, and Peru, constitute a strong representation of the Spanish and Portuguese-speaking world in the selected corpus. This clustering can be related to the choice of keywords, which might be more prevalent in research areas of particular interest in these countries.

Despite the leading role of China and the United States, a notable aspect of our corpus is the wide representation of countries from around the globe, which testifies to the international interest in the research theme. United Kingdom, Germany, Italy, France, Canada, and Australia all feature prominently, reflecting the strong research cultures and infrastructures in these countries. Interestingly, India, known for its robust IT industry and fast-growing academic research sector, seems to be underrepresented, with only 105 publications. This could be related to the research themes being less represented in Indian academic circles, or

perhaps a reflection of the research output in English-language databases, which may not capture the full scope of research activity in this country.

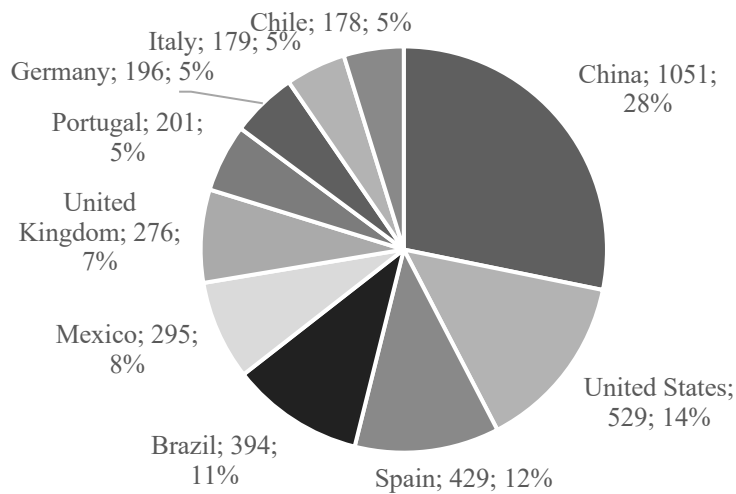


Figure 2. Top 10 Countries in Number of Publications

Table 3. Top 20 Countries in Number of Publications

Country	Publications
China	1051
United States	529
Spain	429
Brazil	394
Mexico	295
United Kingdom	276
Portugal	201
Germany	196
Italy	179
Chile	178
Colombia	148
Peru	148
France	146
Canada	129
Australia	128
India	105
Ecuador	90
Argentina	88
Belgium	78
Iran	65

The list also includes countries with emerging economies such as Ecuador, Argentina, and Iran, highlighting the global relevance and applicability of the selected keywords and their associated research. The wide geographic spread of the contributing authors and the range of countries involved point to a broad, international consensus on the importance of the research topics and highlight the universality of the issues addressed by these keywords.

Such geographical diversity may evidence a variety of research perspectives, enriching the discourse around these topics. Given the significant representation of both developed and emerging economies in the research output, it could be again inferred that these themes have universal relevance, transcending the boundaries of economic and social development. This affirms the importance of continuing global collaboration in exploring these themes in future research.

This collected corpus spans, as shown in figure 3, a broad range of 1587 unique journals and proceedings from 1967 to 2023, and the mean number of publications per journal or proceeding is 1.95, with a median and mode of 1. This distribution indicates a wide range yet a significant portion of them only contains a single related publication, pointing to a possible lack of specialized outlets for this particular theme.

Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) holds the top rank with the highest number of publications, 78 in total. This series is known for its focus on new developments and applications in computer science and information technology, suggesting that the topic is gaining traction within these fields, specifically in Artificial Intelligence and Bioinformatics subdisciplines.

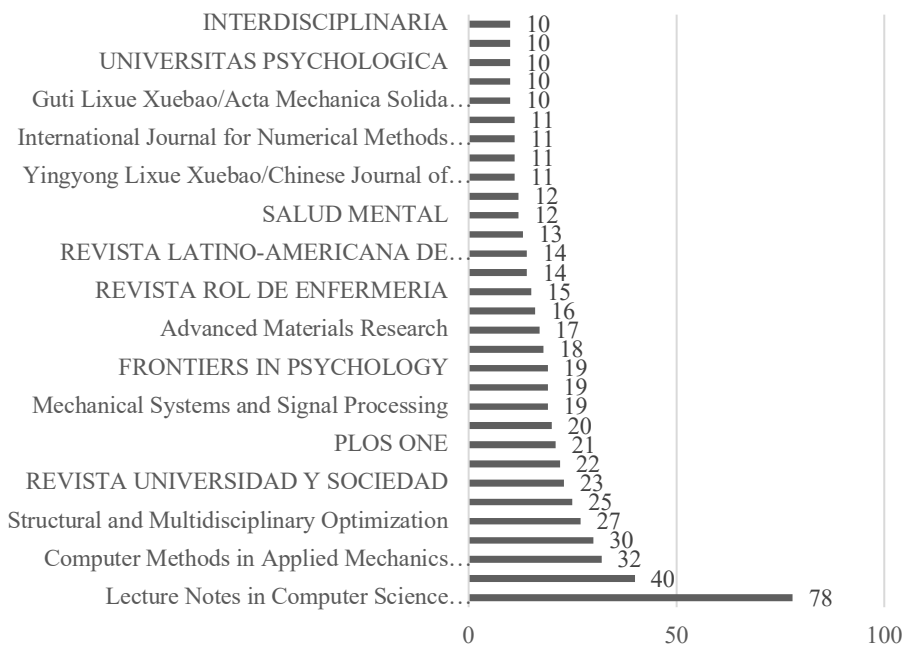


Figure 3. Publications per Journal/Event with 10 or more contributions

The "International Journal of Environmental Research and Public Health" comes next, with 40 publications, and the presence of "Applied Mathematical Modelling" (with 30 publications) and "Structural and Multidisciplinary Optimization" (with 27 publications) on the list indicates the utilization and exploration of non-probabilistic data within the realm of applied mathematics and optimization problems.

It is also notable that journals such as "Retos-Nuevas Tendencias en Educación Física Deporte y Recreación" and "Revista Universidad y Sociedad" appear on the list with 25 and 23 publications, respectively, suggesting that the topic is gaining acceptance and application in social sciences and sports research, emphasizing its interdisciplinary reach.

Among other top publishing sources, "Sustainability" and "PLOS ONE" also appear, with 22 and 21 publications respectively. These are both broad scope journals, and their inclusion here underlines the wide applicability of non-probabilistic methodologies across various research domains.

Lastly, the appearance of titles like "Jisuan Lixue Xuebao/Chinese Journal of Computational Mechanics" and "Jixie Qiangdu/Journal of Mechanical Strength" highlights the global engagement, reflecting a keen interest from different regions and demonstrate wide adoption in computational mechanics and mechanical engineering. It is also coherent with Chine being the country with more publications gathered.

Under the time spam perspective, as presented on figure 4, it can be seen an expressive growth over the past 20 years. During the 44-year period, from 1967 to 2023, the annual number of publications showed considerable variation, with an average of 75.9 publications per year, a median of 7.5, and a mode of 1.

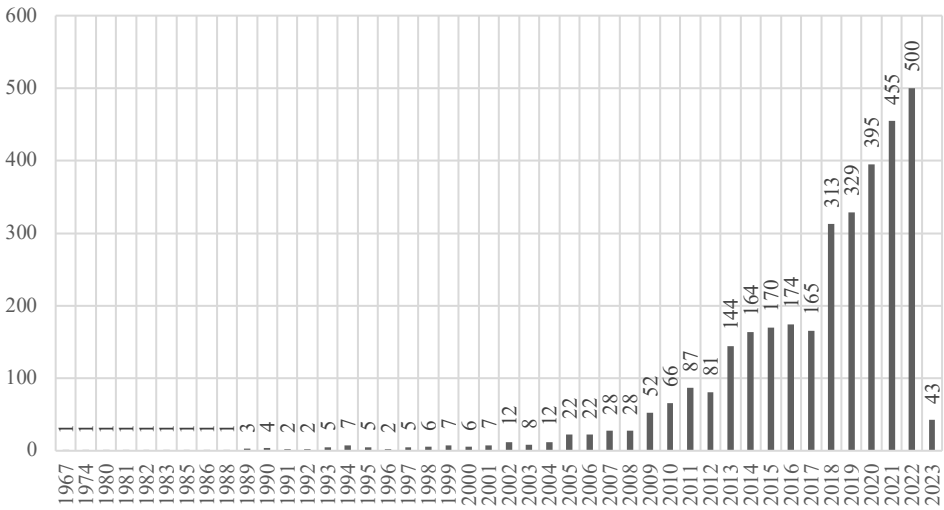


Figure 4. Publications per Year

The number of publications began to increase steadily after 2004, reflecting a growing interest in the field. For instance, the volume of publications more than doubled between 2004 (12 publications) and 2008 (28 publications). The trend continued, albeit at a slightly reduced pace, until 2012, marking a period of consistent growth for the field.

However, between 2012 and 2013, the publication numbers experienced a substantial jump, increasing from 81 to 144. This seems to mark a significant moment, reflecting a broader academic recognition and an uptick in research activity. This upward trend was maintained over the subsequent years, with minor fluctuations in the volume of output.

Notably, the number of publications soared significantly in the last three years, culminating in 2022 with a record-breaking 500 publications. Although a slight reduction is observed in 2023, it is pertinent to remember that the data for this year is incomplete and the final count might be significantly higher.

The cause behind this surge in interest warrants further investigation, and future research should also explore the geographic and institutional distribution of these publications, as well as their thematic focus and the methodologies employed. It's critical to understand whether this trend will continue, as this could signal the maturing as a recognized academic discipline.

The analysis of the number of coauthors per publication, as reproduced on table 4, presents the nature of research collaboration in the field, with a mean of 3.7, and a median and mode both at 3. This suggests that most of the work in this area tends to be collaborative, often involving small to medium-sized teams.

Table 4. Frequency of the number of authors per Publications

Number of Authors	Publications	%
26	1	7.76%
25	1	
24	6	
21	1	
20	2	
19	2	
18	1	
17	2	
16	2	
15	3	
14	2	
13	5	
12	11	
11	15	
10	19	
9	31	
8	58	
7	97	
6	222	37.53%
5	362	
4	669	
3	829	54.71%
2	735	
1	263	

A detailed breakdown reveals that a substantial proportion of the publications (24.83%) had three authors, followed by publications with two authors (22.01%), and four authors

(20.04%). A smaller number of publications (7.88%) had a single author. Single-author publications are less common, which could be due to the increasing complexity of research in this area, possibly requiring various skills that are unlikely to be found in a single researcher. Furthermore, higher number of coauthors, peaking at 26, represents a small fraction of the total (each contributing less than 0.5%).

Out of 9897 unique authors, the average number of publications per author was 1.26, with both the median and mode at 1. This indicates a high level of diversity in authorship, with a majority of single contributions.

Delving deeper into the data, a group of 15 authors emerged with a significantly higher number of contributions, as shown on table 5. These prolific authors, who were a part of multiple research papers, bring forth a diverse array of expertise, as evinced by their varied research backgrounds and fields of specialization.

Table 5. Top 15 most frequent authors

Author	Publications	% of Publications
Lei, Wang	62	1.86%
Xiaojun, Wang	58	1.74%
Zhiping, Qiu	57	1.71%
Yangjun, Luo	17	0.51%
Zhan, Kang	17	0.51%
Yakov, Ben-Haim	15	0.45%
Manuela, Ferreira	15	0.45%
Jie, Liu	14	0.42%
Joao, Duarte	14	0.42%
Chen, Yang	13	0.39%
Yunlong, Li	13	0.39%
Chong, Wang	13	0.39%
David, Moens	13	0.39%
Jiang, C.	13	0.39%
Ruixing, Wang	13	0.39%

Standing out amongst these authors is Lei Wang, with the highest count of 62 publications. Similarly, Xiaojun Wang and Zhiping Qiu have made substantial contributions, with 58 and 57 publications respectively. These authors, point towards the active research environment and the substantial interest in their region.

In addition, Yangjun Luo and Zhan Kang, each with 17 publications, have also contributed notably to the body of literature, with consistent contribution that might reflect a sustained focus on this field of study, and they could potentially be established figures in the domain.

Meanwhile, Yakov Ben-Haim and Manuela Ferreira, both with 15 publications, might bring diverse geographic perspectives. The presence of the first is notable, as he is a prominent researcher in the area of info-gap decision theory, a non-probabilistic uncertainty model, aligning with the topic of non-probabilistic data analysis.

Further, authors such as Jie Liu, Joao Duarte, and David Moens, amongst others, with 13 to 14 publications, again evidence the international reach and multidisciplinary nature of this research field. The contributions from these authors further diversify the knowledge base and approaches.

A deeper level of analysis of the network of coauthorship, presented on figure 5, manifests 35 distinct clusters, enabling to better understand the collaborations and intellectual exchange across the field. Examining the strength of these coauthorship ties, Xiaojun Wang emerges as a central figure with a coauthorship strength of 174, and he belongs to cluster 6. This indicates that Wang's collaborative efforts are not only numerous but also frequently with the same group of researchers, pointing to a consistent and close-knit collaboration, potentially reflecting a stable research team or institutional relationship.

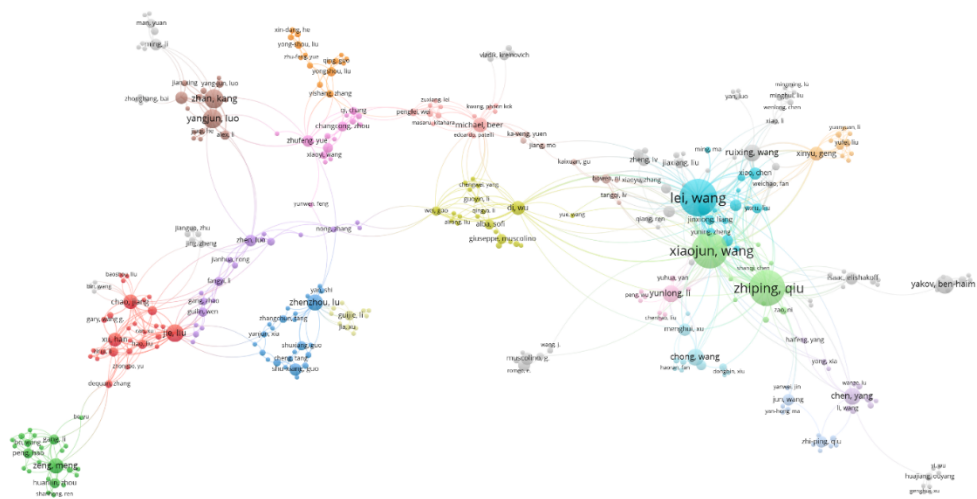


Figure 5. Network View of Co-Authorship

Zhiping Qiu and Jie Liu follow, with coauthorship strength of 163 and 115 respectively, both located in cluster 11. The strong ties between these authors and their colleagues might suggest an established association, or perhaps a mentor-mentee relationship. It is also possible that their shared institutional or geographic locations have facilitated these frequent collaborations.

On another note, Ruixing Wang and Yunlong Li present coauthorship strengths of 51 and 39, respectively, but they are affiliated with different clusters (1 and 26, respectively). This could be indicative of their roles in bridging different research communities, fostering cross-cluster collaborations and potentially facilitating the flow of knowledge and ideas across different groups.

The distribution of authors across various clusters, each with a diverse range of coauthorship strengths, could likely be linked to the multidisciplinary nature of this topic, which invites collaboration from diverse academic backgrounds and geographic locations.

Regarding the average year of publication, a different perspective of the authors' relevance emerges. For instance, the top two authors with the highest scores, Zeshang Li (score: 2021.3333) and Jing Zheng (score: 2020.6667), are associated with clusters 17 and

28, respectively. Notably, while these authors' publication records are more recent on average, their coauthorship strength (10 and 12 respectively) is relatively low compared to some of their counterparts. This could indicate these authors may have more recent entry points into the field, potentially bringing novel perspectives and approaches to their work. Their lower coauthorship strength, meanwhile, could be indicative of a more independent research style or perhaps a broader range of collaborative relationships.

In contrast, researchers such as Yunlong Li and Jie Liu, who show considerable coauthorship strength (39 and 51 respectively), also demonstrate a reasonably contemporary average publication year, with scores of 2019.0769 and 2018.9286 respectively. This intersection of strong collaboration and more recent publication dates may signify their sustained involvement and adaptability within the rapidly evolving field. While Jie Liu is associated with cluster 1, Yunlong Li belongs to cluster 18, indicating the different contexts of their collaborative networks. The diversity in the authors' clusters, coauthorship strengths, and average publication years provides a multi-faceted picture of the research landscape within non-probabilistic data analysis.

A review of the top 20 keywords in terms of co-occurrence strength reveals a blend of themes, ranging from health-related terms like "covid-19", "depression", and "anxiety" to academic-focused ones such as "higher education", "teachers", and "university students". Notably, these co-occurring keywords reflect on the multifaceted implications of the pandemic and its far-reaching impact on global research scenario.

The keyword "covid-19" unsurprisingly leads the list, with a co-occurrence strength of 430, resonating with the global emphasis on recent pandemic research. It is linked to cluster 101, indicating its association with a specific thematic group within the corpus. The high co-occurrence strength implies that covid-19 has been a significant factor influencing research trends.

Mental health-related terms such as "depression", "anxiety", "mental health", "stress", and "quality of life" also demonstrate significant co-occurrence strengths. They are spread across different clusters (96, 15, 90, and 2 respectively), which could suggest that mental health themes are a common thread across various research areas within the corpus. This implies a growing awareness and acknowledgment of mental health issues in diverse research contexts.

Similarly, the prominence of keywords like "higher education", "teachers", and "university students", affiliated to clusters 88, 33, and 27 respectively, indicate the sustained interest in research themes related to education. The high co-occurrence of these terms might be a reflection of the significant changes experienced in the educational sector due to factors like the digital shift triggered by the pandemic.

It's also intriguing to notice more methodological terms such as "reliability", "probability", and "validity", associated with clusters 80 and 97. These terms, integral to research design and measurement, appear with notable co-occurrence strengths, pointing towards the consistent investigative importance of these elements.

Another set of keywords, including "nursing", "health", and "physical activity", connected to clusters 81 and 65, indicates a strong focus on healthcare. This co-occurrence pattern could, again, be interpreted as the impact of the recent health crisis influencing research priorities in the field, or potentially reflecting the importance in health-related studies.

Keywords like "gender", "perception", and "adolescents" (clusters 68, 79, and 91) demonstrate the incorporation of sociodemographic factors and perceptual aspects in related research. The diverse range and their co-occurrence strengths further emphasizes the

multifaceted nature of non-probabilistic data analysis and its wide-ranging applications across different research themes.

Within the top 20 keywords in terms of average year, the majority, like "validation study", "autonomous learning", "beliefs", and "csr", are around 2021. This could suggest that these themes are of recent interest, and potentially represent evolving research directions.

Several of the keywords listed like "engagement", "environmental sustainability", "gender equality", "logistic regression", "resource management", and "theory of planned behavior" point towards the inclusion of societal and environmental aspects, coupled with advanced statistical methods, in the current research discourse. The relatively recent average publication year for these keywords suggests an ongoing shift towards more inclusive and contemporary issues, while reinforcing the relevance of statistical methods like logistic regression in the field.

Certain health-related keywords like "hospitalization", "hypertension", and "mediterranean diet" also have an average publication year of 2021, indicating their recent prominence. This might reflect the influence of the recent health crisis on research themes. On figure 6, the network view of all terms is presented.

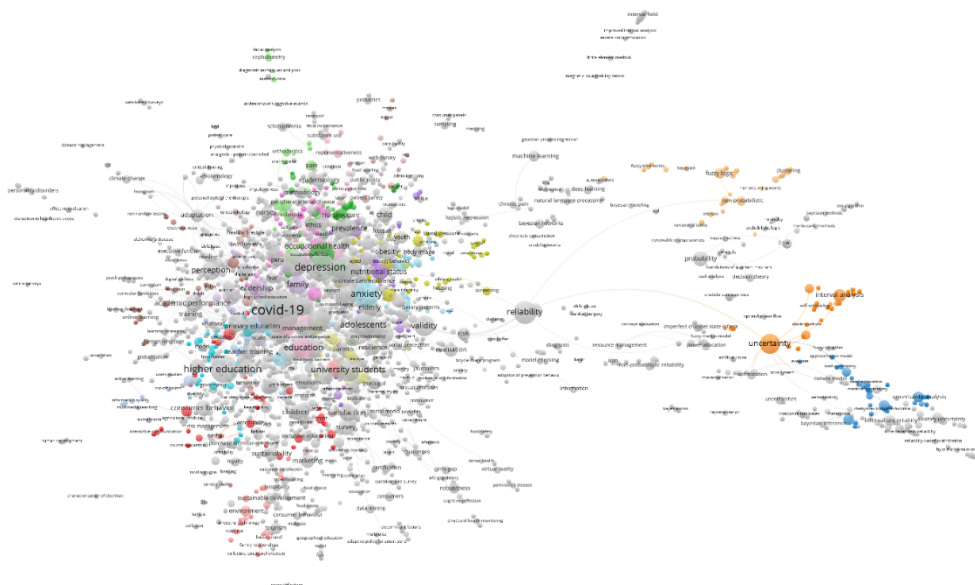


Figure 6. Network View of Co-Occurrence of Terms

Keywords such as "validation study", "autonomous learning", and "professional drivers" could be seen as signifying the broad applicability, reaching into domains of study design, education, and occupational studies. This wide range of keywords with an average publication year of 2021 implies the contemporaneous relevance of non-probabilistic data in diverse research contexts.

The keyword "emotional intelligence" and "benin" have slightly lower average publication years (2021 and 2021, respectively), suggesting that these themes, while still relevant, may have emerged a bit earlier in the research corpus. The range of the 'score' across these keywords showcases the temporal dynamics of research themes within non-

probabilistic data analysis, reflecting how the field continues to evolve and incorporate diverse topics over time.

Examining the citation data of the top 20 most referenced publications, it is evident, as presented on table 6, that specific research themes have a significant impact on the academic discourse. In a general context, papers with the most citations include "Entropy for intuitionistic fuzzy sets" [35], published in Fuzzy Sets and Systems in 2001, and "Antioptimization of structures with large uncertain-but-non-random parameters via interval analysis" [28], published in Computer Methods in Applied Mechanics and Engineering in 1998. It is worth noting that these works date back two decades, yet they still continue to be referenced in recent years, highlighting their foundational contribution to the field.

Table 6. Top 20 Most Cited Publications

Document Title	Authors	Citations
Entropy for intuitionistic fuzzy sets [25]	Szmidt, E; Kacprzyk, J	645
Pan-European survey on the occurrence of selected polar organic persistent pollutants in ground water [29]	Loos, Robert; Locoro, Giovanni; Comero, Sara; Contini, Serafino; Schwesig, David; Werres, Friedrich; Balsaa, Peter; Gans, Oliver; Weiss, Stefan; Blaha, Ludek; Bolchi, Monica; Gawlik, Bernd Manfred	618
Social research 2.0: virtual snowball sampling method using Facebook [3]	Baltar, Fabiola; Brunet, Ignasi	560
Sample selection bias and Heckman models in strategic management research [8]	Certo S.T., Busenbark J.R., Woo H.-S., Semadeni M.	409
Antioptimization of structures with large uncertain-but-non-random parameters via interval analysis [28]	Qiu, ZP; Elishakoff, I	352
A survey of non-probabilistic uncertainty treatment in finite element analysis [25]	Moens, D; Vandepitte, D	318
Photovoltaic solar energy: Conceptual framework [39]	Vasconcelos Sampaio, Priscila Goncalves; Aguirre Gonzalez, Mario Orestes	305
A nonprobabilistic concept of reliability [6]	Benhaim, Y	290
Imprecise probabilities in engineering analyses [5]	Beer, Michael; Ferson, Scott; Kreinovich, Vladik	282
Dealing with uncertainty in model updating for damage assessment: A review [34]	Simoen, Ellen; De Roeck, Guido; Lombaert, Geert	259
Lean manufacturing, non-financial performance measures, and financial performance [14]	Fullerton, Rosemary R.; Wempe, William F.	248

Prevalence of thyroid disorders in the working population of Germany: Ultrasonography screening in 96,278 unselected employees [30]	Reiners, C; Wegscheider, K; Schicha, H; Theissen, P; Vaupel, R; Wrbitzky, R; Schumm-Draeger, PM	230
A review of uncertainty analysis in building energy assessment [36]	Tian, Wei; Heo, Yeonsook; de Wilde, Pieter; Li, Zhanyong; Yan, Da; Park, Cheol Soo; Feng, Xiaohang; Augenbroe, Godfried	217
Correlation analysis of non-probabilistic convex model and corresponding structural reliability technique [15]	Jiang, C.; Han, X.; Lu, G. Y.; Liu, J.; Zhang, Z.; Bai, Y. C.	216
Comparison of dynamic response of structures with uncertain-but-bounded parameters using non-probabilistic interval analysis method and probabilistic approach [29]	Qiu, ZP; Wang, XJ	201
Joint User Scheduling and Power Allocation Optimization for Energy-Efficient NOMA Systems With Imperfect CSI [13]	Fang, Fang; Zhang, Haijun; Cheng, Julian; Roy, Sebastien; Leung, Victor C. M.	194
Continuum topology optimization with non-probabilistic reliability constraints based on multi-ellipsoid convex model [22]	Luo, Yangjun; Kang, Zhan; Luo, Zhen; Li, Alex	192
Engineering computation under uncertainty - Capabilities of non-traditional models [23]	Moeller, Bernd; Beer, Michael	191
Non-probabilistic finite element analysis for parametric uncertainty treatment in applied mechanics: Recent advances [24]	Moens, David; Hanss, Michael	188
Quantum theory of probability and decisions [10]	Deutsch, D	187

A deeper look on year-by-year citation distribution from 2019 to 2023 shown on figure 7 reveals a diverse range of themes, with the study "Social research 2.0: virtual snowball sampling method using Facebook" [3], published in Internet Research, cumulating a considerable number of citations. This indicates the relevance and applicability of non-probabilistic data analysis methods in the field of social research, specifically in the era of social media and digital connectivity. Furthermore, it also implies the growing interest in online research methods, which often rely on non-random samples due to the nature of the internet.

In another context, the publication "Sample selection bias and Heckman models in strategic management research" [8] in the Strategic Management Journal in 2016, focuses on the impact of sample selection bias, another central aspect in non-probabilistic data analysis.

This reflects the increasing concern about bias and its impact on the reliability and validity of research outcomes, particularly in strategic management studies.

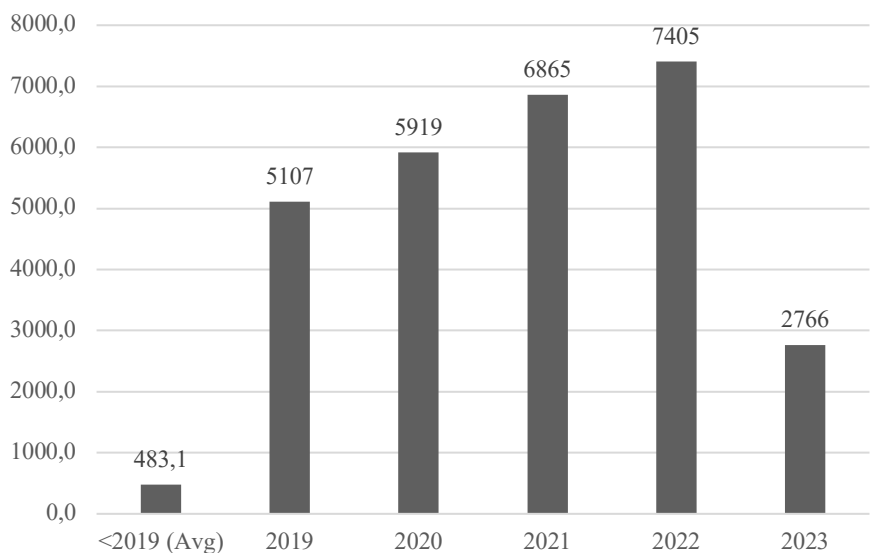


Figure 7. Number of Citations per Year

Interestingly, the citation patterns also indicate the interdisciplinary nature of non-probabilistic data analysis applications. The publications “Pan-European survey on the occurrence of selected polar organic persistent pollutants in ground water” [21] in Water Research and “Photovoltaic solar energy: Conceptual framework” by [39] in Renewable & Sustainable Energy Reviews, are cases in point. These studies demonstrate the importance of non-probabilistic methods in environmental and energy-related research.

A noteworthy observation is that as the total citation count reaches over ten thousand, the resulting h-index stands at 78, indicating a high impact and relevance of the corpus in the broader research context. Further exploration of citation patterns and trends can uncover more nuanced insights about the evolution and future directions of non-probabilistic data analysis.

4. Final considerations

This study offers a comprehensive bibliometric analysis on the topic of non-probabilistic data, with the aim of illuminating the intellectual structure and evolution of the field. The examination of co-authorship networks and the assessment of keyword co-occurrence provided an understanding of the underlying social and thematic patterns prevalent in this research area [48]. The contemporary nature of the researchers was gauged, revealing the key individuals contributing to the recent scholarly discourse.

The co-authorship analysis highlighted the collaborative nature while the emergence of robust clusters pointed towards the existence of well-established research communities. The varying degrees of co-authorship strength indicated the dynamic nature of those collaborations, reflecting the ongoing shifts in the research partnerships and alliances [27]. Furthermore, the analysis of the publication years emphasized the recent contributions, spotlighting the researchers who are pushing the boundaries of the field.

The keyword co-occurrence analysis exposed the pivotal themes underpinning non-probabilistic data research. The identified clusters, formed based on the frequency of keyword co-occurrence, underscored the major thematic areas that have shaped the research landscape in this field. The strength of co-occurrence further delineated the predominant trends and current research focus, signifying the centrality of themes like "covid-19", "depression", "anxiety", and "higher education" in recent literature [9].

The study also underscored contemporary keywords such as "validation study", "autonomous learning", "beliefs", and "engagement", suggesting the alignment with modern research foci such as validation of data analysis methods, learner autonomy in education, belief systems in social research, and engagement in the era of digital communication. From the corpus analysed, it is observed that recent publications have also dedicated to investigating and proposing new methods to analyse non-probabilistic data and to provide quality metrics [33][41][45][47], what reinforces the increased highlight being given to this type of data.

However, while this study provides a structured analysis of the field, it is important to note the inherent limitations. The search for literature was confined to the Web of Science (WoS) and Scopus databases, potentially excluding relevant research published elsewhere. Future studies may expand the database sources to capture a broader spectrum of non-probabilistic data research.

Future research might also look into a deeper exploration of the thematic clusters within non-probabilistic data research. A more qualitative approach could supplement the findings of this study, providing an in-depth understanding of the content and context of the research themes. Additionally, tracking the evolution of co-authorship networks and keyword co-occurrence over time can provide valuable insights into the dynamic changes in the field.

The insights gleaned from this bibliometric analysis offer a structured understanding of the field, while also revealing potential avenues for future research. The application of bibliometric methods in this manner represents a valuable tool in scientific analysis, allowing for an objective review of the research field [12].

References

- [1] Aven, T. (2013). On how to define, understand and describe risk. *Reliability Engineering & System Safety*, 144, 1-10.
- [2] Aven, T. (2016). Risk assessment and risk management: Review of recent advances on their foundation. *European Journal of Operational Research*, 253(1), 1-13.
- [3] Baltar, F., & Brunet, I. (2012). Social research 2.0: virtual snowball sampling method using Facebook. *Internet Research*.

- [4] Battaglia, A. G. (2008). Non-probabilistic treatment of uncertainty in engineering design: Literature review and application to composites. *Composites Part B: Engineering*, 39(5), 750-756.
- [5] Beer, M., Ferson, S., & Kreinovich, V. (2013). Imprecise probabilities in engineering analyses. *Mechanical Systems and Signal Processing*.
- [6] Ben-Haim, Y. (1994). A nonprobabilistic concept of reliability. *Structural Safety*.
- [7] Borgatti, S. P. (2005). Centrality and network flow. *Social networks*, 27(1), 55-71.
- [8] Certo, S. T., Busenbark, J. R., & Woo, H.-S. (2016). Sample selection bias and Heckman models in strategic management research. *Strategic Management Journal*.
- [9] Cobo, M. J., López-Herrera, A. G., Herrera-Viedma, E., & Herrera, F. (2011). Science mapping software tools: Review, analysis, and cooperative study among tools. *Journal of the American Society for Information Science and Technology*, 62(7), 1382-1402.
- [10] Deutsch, D. (1999). Quantum theory of probability and decisions. *Proceedings of the Royal Society A-Mathematical Physical and Engineering Sciences*, 455(1988), 3129-3137.
- [11] Dubois, D., & Prade, H. (1988). Possibility theory: An approach to computerized processing of uncertainty. Plenum Press.
- [12] Ellegaard, O., & Wallin, J. A. (2015). The bibliometric analysis of scholarly production: How great is the impact?. *Scientometrics*, 105(3), 1809-1831.
- [13] Fang, F., Zhang, H., Cheng, J., Roy, S., & Leung, V. C. M. (2017). Joint User Scheduling and Power Allocation Optimization for Energy-Efficient NOMA Systems With Imperfect CSI. *IEEE Journal on Selected Areas in Communications*, 35(12), 2874-2885.
- [14] Fullerton, R. R., & Wempe, W. F. (2009). Lean manufacturing, non-financial performance measures, and financial performance. *International Journal of Operations & Production Management*.
- [15] Jiang, C., Han, X., Lu, G. Y., Liu, J., Zhang, Z., & Bai, Y. C. (2011). Correlation analysis of non-probabilistic convex model and corresponding structural reliability technique. *Computer Methods in Applied Mechanics and Engineering*.
- [16] Klir, G. J., & Yuan, B. (1995). Fuzzy sets and fuzzy logic: Theory and applications. Prentice Hall.
- [17] Kosko, B. (1992). Neural networks and fuzzy systems: A dynamical systems approach to machine intelligence. Prentice Hall.
- [18] Liu, B., & Liu, Y. K. (2002). Expected value of fuzzy variable and fuzzy expected value models. *IEEE Transactions on Fuzzy Systems*, 10(4), 445-450.
- [19] Liu, H. C., & Pedrycz, W. (2018). Evaluation of feature selection approaches in software defect prediction. *Information Sciences*, 451-452, 135-147.

-
- [20] Liu, H. C., You, J. X., You, X. Y., & Shan, M. M. (2015). A novel approach for failure mode and effects analysis using combination weighting and fuzzy VIKOR method. *Applied Soft Computing*, 28, 579-588.
- [21] Loos, R., Locoro, G., Comero, S., Contini, S., Schwesig, D., Werres, F., Balsaa, P., Gans, O., Weiss, S., Blaha, L., Bolchi, M., & Gawlik, B. M. (2010). Pan-European survey on the occurrence of selected polar organic persistent pollutants in ground water. *Water Research*.
- [22] Luo, Y., Kang, Z., Luo, Z., & Li, A. (2009). Continuum topology optimization with non-probabilistic reliability constraints based on multi-ellipsoid convex model. *Structural and Multidisciplinary Optimization*, 39(3), 273-290.
- [23] Moeller, B., & Beer, M. (2008). Engineering computation under uncertainty-capabilities of non-traditional models. *Computers & Structures*, 86(10), 1044-1061.
- [24] Moens, D., & Hanss, M. (2011). Non-probabilistic finite element analysis for parametric uncertainty treatment in applied mechanics: Recent advances. *Finite Elements in Analysis and Design*, 47(1), 4-16.
- [25] Moens, D., & Vandepitte, D. (2005). A survey of non-probabilistic uncertainty treatment in finite element analysis. *Computer Methods in Applied Mechanics and Engineering*.
- [26] Mongeon, P., & Paul-Hus, A. (2016). The journal coverage of Web of Science and Scopus: a comparative analysis. *Scientometrics*, 106(1), 213-228.
- [27] Newman, M. E. (2001). The structure of scientific collaboration networks. *Proceedings of the national academy of sciences*, 98(2), 404-409.
- [28] Qiu, Z. P., & Elishakoff, I. (1998). Antioptimization of structures with large uncertain-but-non-random parameters via interval analysis. *Computer Methods in Applied Mechanics and Engineering*.
- [29] Qiu, Z. P., & Wang, X. J. (2003). Comparison of dynamic response of structures with uncertain-but-bounded parameters using non-probabilistic interval analysis method and probabilistic approach. *International Journal*.
- [30] Reiners, C., Wegscheider, K., Schicha, H., Theissen, P., Vaupel, R., Wrbitzky, R., & Schumm-Draeger, P. M. (2004). Prevalence of thyroid disorders in the working population of Germany: Ultrasonography screening in 96,278 unselected employees. *Thyroid*.
- [31] Sartori, J.T.D. (2023). True Confidence Level of Real-World Data: implications of non-probabilistic data on decision-making and research. *Italy*, 48p.
- [32] Shafer, G. (1976). *A mathematical theory of evidence* (Vol. 1). Princeton university press.
- [33] Silver Ian, A. & Kelsay James, D. (2023) The moderating effects of population characteristics: a potential biasing factor when employing non-random samples to conduct experimental research. *Journal of Experimental Criminology*, 19(1), 107-118.
- [34] Simoen, E., De Roeck, G., & Lombaert, G. (2015). Dealing with uncertainty in model updating for damage assessment: A review. *Mechanical Systems and Signal Processing*.

- [35] Szmidt, E., & Kacprzyk, J. (2001). Entropy for intuitionistic fuzzy sets. *Fuzzy Sets and Systems*.
- [36] Tian, W., Heo, Y., de Wilde, P., Li, Z., Yan, D., Park, C. S., Feng, X., & Augenbroe, G. (2018). A review of uncertainty analysis in building energy assessment. *Renewable & Sustainable Energy Reviews*.
- [37] Trochim, W. M. (2006). The research methods knowledge base (Vol. 2). Atomic Dog.
- [38] Van Eck, N. J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, 84(2), 523-538.
- [39] Vasconcelos Sampaio, P. G., & Aguirre Gonzalez, M. O. (2017). Photovoltaic solar energy: Conceptual framework. *Renewable & Sustainable Energy Reviews*.
- [40] Wang, Y. M., & Guan, Z. L. (2018). Environmental performance evaluation using a non-radial DEA model: In case of lake water quality in China. *Journal of Cleaner Production*, 176, 63-74.
- [41] Yi-Chuan, F., Yong-Juan, W., Jin-Long, S. & Tong-Guang, G. (2023) Reliability analysis of mechanisms with mixed uncertainties using polynomial chaos expansion. *Quality and Reliability Engineering International*, 39(4), 1248-1268.
- [42] Zadeh, L. A. (1978). Fuzzy sets as a basis for a theory of possibility. *Fuzzy sets and Systems*, 1(1), 3-28.
- [43] Zadeh, L. A. (1988). Fuzzy logic. *Computer*, 21(4), 83-93.
- [44] Zadeh, L. A. (1996). Fuzzy sets. In *Fuzzy Sets, Fuzzy Logic, And Fuzzy Systems: Selected Papers by Lotfi A Zadeh* (pp. 394-432). World Scientific.
- [45] Zeng, M., Changquan, L. & Peng, H. (2023) Unified reliability-based design optimization with probabilistic - uncertain-but-bounded and fuzzy variables. *Computer Methods in Applied Mechanics and Engineering*, 407.
- [46] Zhang, D., & Jiang, B. (2019). Bibliometric and visualized analysis of emergy research. *Ecological Modelling*, 392, 78-88.
- [47] Zhaoping, T., Jun, P., Jianping, S. & Xin, M. (2022) Non-Probabilistic Reliability Analysis of Robot Accuracy under Uncertain Joint Clearance. *Machines*, 10(10).
- [48] Zupic, I., & Čater, T. (2015). Bibliometric methods in management and organization. *Organizational Research Methods*, 18(3), 429-472.