

Machine Learning Applied to Water Distribution Networks Issues: A Bibliometric Review

H. DENAKPO^{1,*}, P. HOUNGUE², T. DAGBA³ and J. DEGILA⁴

¹Institut de Mathématiques et de Sciences Physiques (IMSP), Dangbo, Bénin

²Institut de Mathématiques et de Sciences Physiques (IMSP), Dangbo, Bénin

³Ecole Nationale d'Economie Appliquée et de Management (ENEAM), Cotonou, Bénin

⁴Institut de Mathématiques et de Sciences Physiques (IMSP), Dangbo, Bénin

Abstract

INTRODUCTION: Water Distribution Networks are critical infrastructures that have garnered increasing interest from researchers.

OBJECTIVES: This article conducts a bibliometric analysis to examine trends, the geographical distribution of researchers, hot topics, and international cooperation in using Machine Learning for Water Distribution Networks over the past decade.

METHODS: Using “water distribution” AND (prediction OR “Machine learning” OR “ML” OR detection OR simulation), as search string, 4859 relevant publications have been retrieved from WoS database. After applying the PRISMA method, we retained 2427 documents for analysis with a Bibliometric library programmed in R.

RESULTS: China and the USA are the most productive on the ground, and only one African country appears in this ranking in 14th place. We also identified two ways for future research works, which are: the assessment of water quality and the design of optimisation models.

CONCLUSION: The application of this research in African countries would be fascinating for a better quality of service and efficient management of this resource, which is inaccessible to many African countries.

Keywords: Water Distribution Networks, Machine Learning, Bibliometric study, Bibliometric

Received on 19 December 2023, accepted on 20 March 2024, published on 27 March 2024

Copyright © 2024 H. Denakpo *et al.*, licensed to EAI. This is an open access article distributed under the terms of the [CC BY-NC-SA 4.0](#), which permits copying, redistributing, remixing, transformation, and building upon the material in any medium so long as the original work is properly cited.

doi: 10.4108/ew.5567

*Corresponding author. Email: hermione.denakpo@imsp-uac.org

1. Introduction

Planet Earth is 71% covered in water, and 96,5% is salt water [1]. Moreover, we also have sources of drinking water such as lakes and groundwater, which brings the quantity of fresh water available on earth to 2.5% [1], which makes this resource a rare commodity. In this context of scarcity of resources, public and private water operators are set up to manage and efficiently distribute drinking water to the population because it represents one of the most basic human needs [2]. The main objective of the Water Distribution Network (WDN) is to provide the population with drinking water needs whenever the need arises.

However, these operators face growing and complex challenges such as monitoring water quality, reducing water losses, reducing capital and operating costs, and ensuring network and energy performance [3].

Several solutions have been implemented over time, such as leakage detection and localisation [4], [5], [6], demand forecasting [7], sensors placement for water quality [8], [9], optimisation of WDN as well as the modeling of the interdependence with other infrastructure such as electrical networks [10].

According to the growing work, we are wondering: What is the ongoing research in water distribution networks? What are the main lines of investigation? What is future research? To answer these questions, several reviews have

been made to guide future work in several fields, such as the control of water leaks [11], methods of placing sensors for the detection of contamination [12], and substitution modeling based on Machine Learning [13]. We conduct a bibliometric review to understand better the evolution of using Machine Learning in water distribution networks.

The main axes covered in this paper are the research methodology presented in Section 2 and the highlights of our results and discussions in Section 3. Finally, Section 4 addresses the limits and Section 5 concludes with some research orientations.

2. Methodology

This article presents a bibliometric review based on extracted data from the Web of Science (WoS) literature search database designed to support scientific and scholarly research [14]. The bibliometric data was extracted, on April 28, 2023, from the WoS database using the following search string: “water distribution” AND (prediction OR “Machine learning” OR “ML” OR detection OR simulation). The search identified 4859 relevant publications using Machine Learning (ML) in Water Distribution Networks (WDN) for further analysis. Those publications include articles, Book chapters, early access articles, books, editorial material, reviews and proceedings papers. The selected records were exported to BibTeX format files from WoS. **Figure 1** shows the selection process scheme.

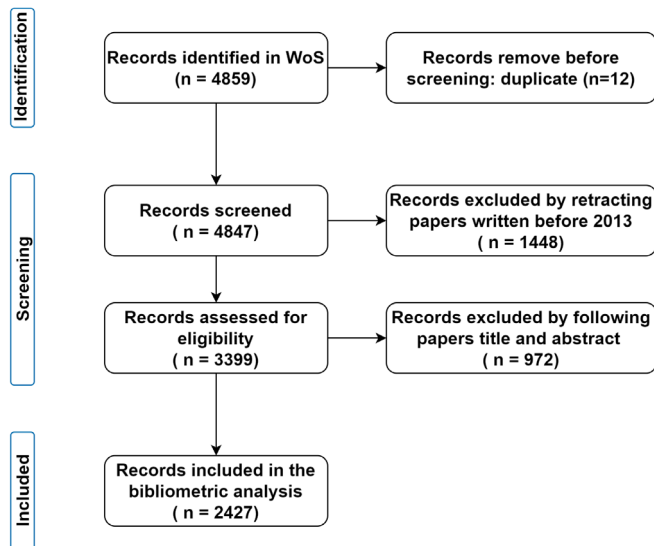


Figure 1. Documents selection process

In summary, for a more in-depth analysis, 2427 relevant publications from the last ten years, from 2013 to April 28, 2023, have been identified from the publications presented in WoS. Thus, the relevant documents acquired are effectively used to perform bibliometric analysis using biblioshiny, an interface of the Bibliometrix library in R.

3. Results and discussion

This section introduces and discusses the main results from our bibliometric study. We give, among other things, an overview of the types of information collected, the most relevant authors, the most cited documents, the conceptual structure map, and the country collaboration.

3.1. Main documents on the use of ML in WDN

We collected various information with 4859 records from 1975 to 2023. We notice that, from 1975 to 1990, very few articles in the field were produced, with a minimum of 0 and a maximum of 3 per year. Therefore, we have chosen 2013 to 2023 to gather more accurate and relevant publications. In addition, other projections were made, as shown in section 2. A total of 2427 publications were completed by 5055 authors, of which 89 single authors wrote 89. International collaboration is about 27.28%. Publications in this study include 71.77% of articles and 23% of proceedings papers. On average, the documents are less than five years old. This means that much of the work has been done recently on this topic. **Table 1** shows the information data collected from the WoS database containing publications related to the use of ML in Water Distribution Networks.

Table 1. Main information about documents on ML in WDN from WoS

Description	Results
Main Information About The Data	
Timespan	2013:2023
Sources (Journals, Books, etc.)	812
Documents	2427
Annual Growth Rate %	-6.49
Document Average Age	4.74
Average citations per doc	11.32
References	48094
Document Contents	
Keywords Plus (ID)	2386
Author's Keywords (DE)	5706
AUTHORS	
Authors	5055
Authors of single-authored docs	69
Authors Collaboration	
Single-authored docs	89
Co-Authors per Doc	3.69
International co-authorships %	27.28
Document Types	
Article	1742
Article; book chapter	13
Article; early access	24
Article; proceedings paper	15
Book	2
Correction	1
Editorial material	10
Proceedings paper	561
Review	59

3.2. Annual scientific production and key sources annual scientific production

Annual Scientific Production shows the evolution of scientific production over the years. The scientific output per year on using ML in WDN from 2013 to 2023 is presented in **Figure 2**. From 2013 to 2014, we noticed a growing interest in the field, which decreased slightly and stagnated in 2015 and 2016. Then, from 2017 to 2022, we saw an increasing evolution of the work carried out in the field. A slight decrease was observed in 2019, which can be explained by the beginning of the global pandemic of Covid 19, which mobilised a large part of scientists and public attention [15]. Regarding the low rate observed in 2023, the information was retrieved from the WoS database at the beginning of the 2nd quarter of 2023 with 24 articles in early access, which suggests that other publications will be made during the year. This demonstrates the growing interest of scientists in the subject.

The top 20 most relevant sources on using ML in WDN are shown in **Table 2**. At the top of the list, we have the “Journal of Water Resources Planning and Management” with 222 publications. It is not surprising in the sense that the publications of this journal address, among others, topics related to innovative technologies, applications, and emerging systems analysis practices to improve the monitoring, modeling, digitisation, and management of water resources [16]. It is followed by the “Water” journal, an open access international and interdisciplinary journal covering all aspects of water with 141 publications, “Water Resources Management” journal with 97 publications, “Water Research” with 61 publications and in fifth place, “Journal Of Hydroinformatics” with 56 publications.

3.3. Source growth dynamics

Source growth dynamics allows us to identify the most relevant sources and assess the progress of scientific production. It appears in **Figure 3** that the first five relevant sources on ML in WDN, depending on the number of documents published, have grown dynamically. The “Water” journal had no publication in the field before 2015 but has exceeded most other journals. It can be explained by the fact that this journal is free to access, it covers all areas related to water, and there is no restriction on the number of pages of papers published [17].

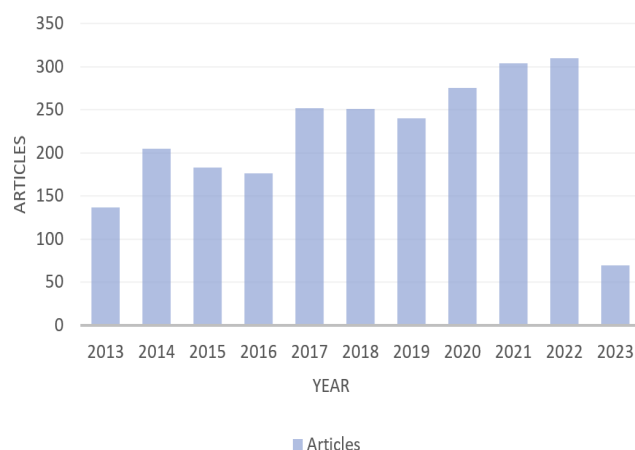


Figure 1. Annual Scientific Production

Table 1. Top 20 most relevant sources

Sources	Articles
“Journal Of Water Resources Planning And Management”	222
“Water”	141
“Water Resources Management”	97
“Water Research”	61
“Journal Of Hydroinformatics”	56
“Urban Water Journal”	54
“Journal Of Pipeline Systems Engineering And Practice”	36
“12 th International Conference On Computing And Control For The Water Industry, Ccwi2013”	33
“Ifac Papersonline”	33
“Water Science And Technology-Water Supply”	33
“Computing And Control For The Water Industry (Ccw2015): Sharing The Best Practice In Water Management”	32
“16 th Water Distribution System Analysis Conference (Wdsa2014) : Urban Water Hydroinformatics And Strategic Planning”	31
“Water Supply”	31
“Journal Of Hydraulic Engineering”	25
“Journal Of Water Supply Research And Technology-Aqua”	24
“Sensors”	22
“Xviii International Conference On Water Distribution Systems, Wdsa2016”	21
“Environmental Modelling \& Software”	20
“Science Of The Total Environment”	20
“Water Resources Research”	19

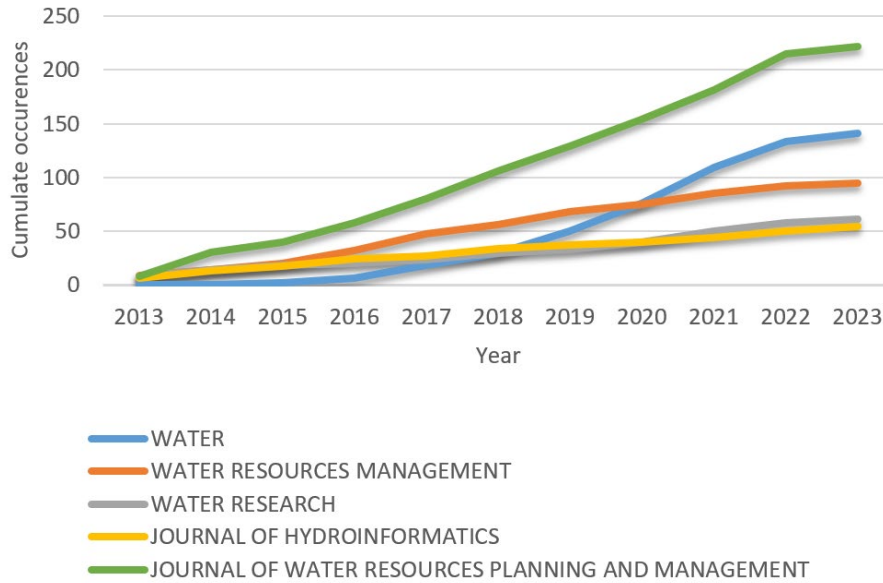


Figure 2. Source Growth Dynamics

3.4. Top 20 pertinent authors based on the number of publications

Figure 4 presents the twenty (20) pertinent authors with the most publications. All top 10 authors are from the academic sector, and the top 3 have over 30 publications to their credit; Avi Ostfeld leads with 59 publications. He is a member of the Technion-Israel Institute of Technology permanent committee for tenure appointments and the present Head of the undergraduate Technion exposure program on hydrology and water resources [18]. We also noticed that among the top 10, three authors were Italian, of which two were from the University of Bari, followed by Australia and Spain with 2 authors from each country.

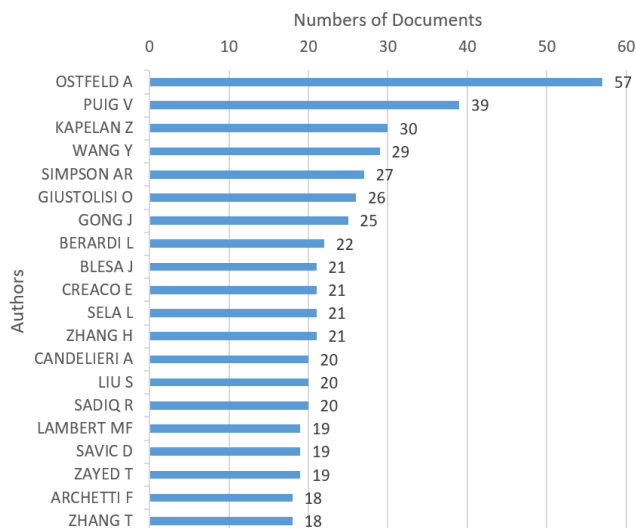


Figure 3. Top 20 Relevant Authors

3.5. Top 20 most pertinent affiliations

Figure 5 presents the most pertinent affiliations on the studied topic from the WoS database. Zhejiang University (China) first has 126 papers and the University of Adelaide (Australia) comes close behind with 123 publications. The University of Exeter (England) holds third place, with 107 publications. Technion University (Israel) is in fourth place with 101 publications, followed by Tongji University (China) with 82 articles. All affiliations are universities from America, Oceania, Asia, and Europe, but none from Africa. We also noticed that four of the top 20 universities are from China.

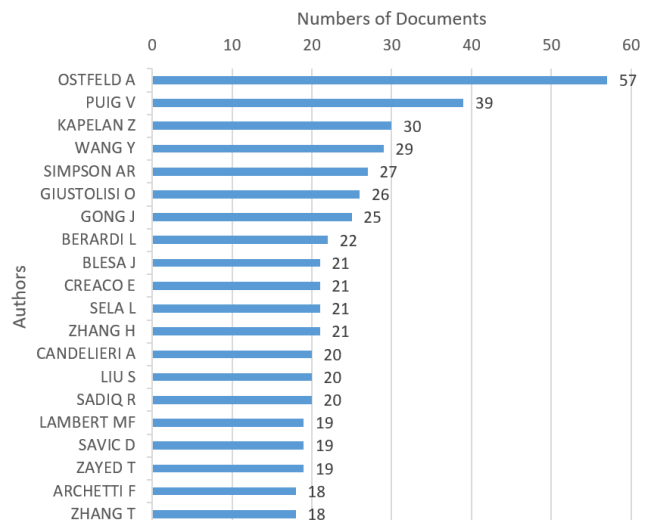


Figure 4. Top 20 pertinent affiliations

3.6. Top 20 most cited documents

Figure 6 presents the most cited documents in the field from 2013 to 2023. First, the “International Conference on Artificial Neural Networks (ICANN, 2019)” introduced the main themes of “Brain-Inspired Computing” and “Machine Learning Research”. It covered all major research areas dealing with neural networks [19], which was cited 243 times. Moreover, we have “A Low-Cost Sensor Network for Real-Time Monitoring and Contamination Detection in Drinking Water Distribution Systems” presented by Theofanis P. Lambrou and al in 2014, cited 153 times [20]. Finally, “Evaluating the risk of water mains failure using a Bayesian belief network model” published by Kabir and al. in 2015, is cited 148 times. The article presents the risk of failure of metallic water pipes using structural integrity, hydraulic capacity, water quality, and consequence factors based on the Bayesian Belief Network (BBN) model [21].

The fourth and fifth places, we have respectively “Novel Leakage Detection by Ensemble CNN-SVM and Graph-Based Localization in Water Distribution Systems” by Jiheon Kang and al. in 2018 [5], cited 147 times, and “Modeling the resilience of critical infrastructure: the role of network dependencies”, by Roberto Guidotti and al in 2016 [10], which is cited 139 times.

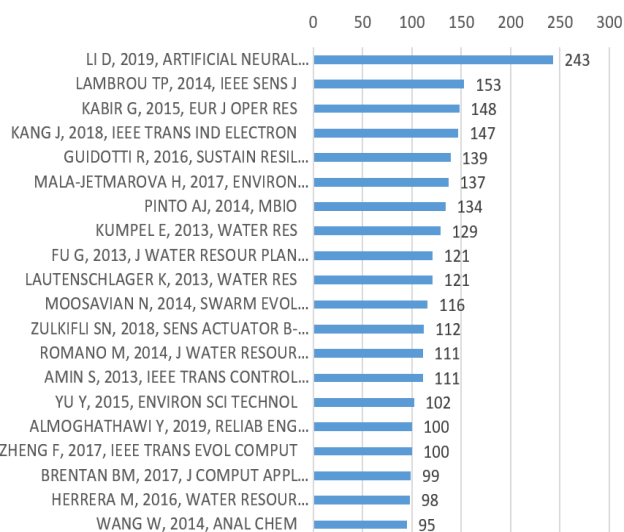


Figure 5. Top 20 most cited documents

3.7. Top 20 scientific production on the use of ML in Water Distribution Networks

Scientific production by country on Machine Learning in Water Distribution Networks is presented in **Table 3**. China occupies 1st place with 1077 documents followed by the United States in second place with 950 records. Third place is occupied by Italy with 639 documents and in fourth place we have the United Kingdom with 461 documents. Other countries on this shortlist produced less than 400 documents. Overall, developed countries have more production.

Moreover, Asia and Europe dominate the ranking. However, one country from Africa (South Africa) ranks 14th with 147 records.

Table 2. Top 20 Scientific Production Countries

Rank	Region	Frequency
1	China	1077
2	Usa	950
3	Italy	639
4	Uk	461
5	Iran	381
6	Australia	351
7	Canada	349
8	India	338
9	Spain	326
10	South Korea	212
11	Brazil	181
12	Netherlands	179
13	Poland	170
14	South Africa	147
15	Israel	146
16	Germany	142
17	France	137
18	Portugal	126
19	Mexico	112
20	Singapore	92

As shown in **Figure 7**, until 2021, the USA was the most productive country. However, an important grow in publications in China from 2022 is noticed, surpassing the 950 publication marks. This can be justified because COVID-19 has highlighted the need to monitor better infrastructures as critical as water distribution networks, which can be an axis of propagation of several contaminants that can exacerbate COVID-19 symptoms [22]. As a reminder, China was the starting point of the pandemic in 2019 and one of the countries most affected by COVID-19 [23].

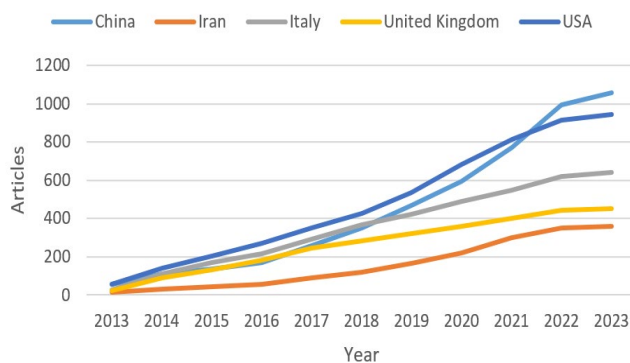


Figure 6. Country Production over Time

3.8. Words Cloud Related to the Use of ML in Water Distribution Networks

Figure 8 and Table 4 present the words Cloud and the most cited word Cloud, related to the use of ML in WDNs. We used “keywords plus” for this bibliometric analysis.

Furthermore, we have taken care to remove from this analysis the terms duplicated by being plural, such as Model-Models, Algorithm-Algorithms, or systems-system, to get an accurate analysis.



Figure 7. Words Cloud

The first three most cited terms are “Model”, with an occurrence of 354; “Water Distribution Systems,” with 281 occurrences and “Design”, with 235 occurrences. Thus, the latest applications of Machine Learning are water distribution systems design and the application of different simulation models. Designing and optimizing water distribution systems is crucial to managing water distribution networks. Various simulation models and algorithms, including genetic algorithms, are used to improve reliability, quality, and water demand prediction. Prediction models and sensor networks also detect pipeline leaks and evaluate the distributed water quality. The terms in Table 4 illustrate the main challenges WDNs face and the focus of much research.

Terms	Frequency
Water Distribution Networks	128
Networks	119
Location	102
Drinking-Water	101
Pipelines	98
Reliability	97
Quality	93
Demand	86
Prediction	86
Leak Detection	73
Genetic Algorithms	64
Water	62

Table 3. 20 Most cited words

Terms	Frequency
Model	354
Water Distribution-Systems	280
Design	235
Systems	212
Optimisation	202
Simulation	182
Management	143
Algorithm	130

3.9. Co-citation network

Co-citation relationships are established when other documents cite two documents together. A Co-citation Network between 50 authors in Figure 9 shows that 3 clusters are observed around Rossman L.A. (2000), Ostfeld A. (2008), and Puust R. (2010). They worked respectively on “Epanet 2 user’s manual” [24], an “ant colony optimisation methodology for the cost-effective design of gravitational water distribution systems” [25], and a “review of methods for leakage management in pipe networks” [26].

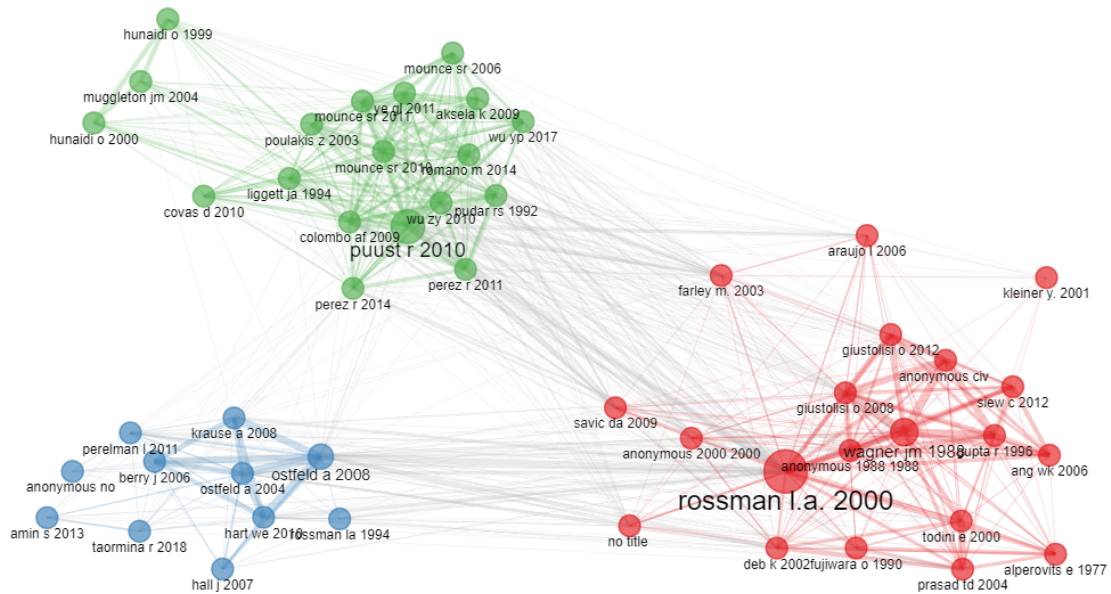


Figure 8. Co-citation Networks

analysis method: two clusters were identified. The words of the first cluster are around “identification”, “placement”, “prediction,” and “systems”. In the second cluster, the terms are more around “drinking water” and “disinfection”.

3.10. Conceptual structure map

The map of conceptual structure is presented in **Figure 10**. We used “keywords plus” and multiple correspondence

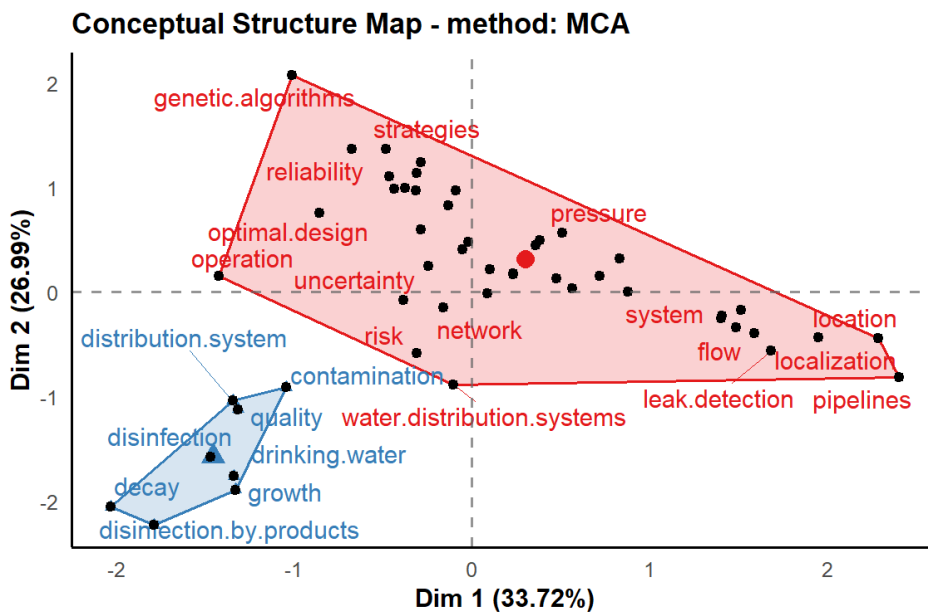


Figure 9. Conceptual structure map

3.11. Co-occurrence Network

Co-occurrence network analysis, namely the way to highlight the relationship between words of 50 “keywords

plus” in **Figure 11**, allows us to identify 4 clusters of words around “models”, “water distribution systems”, “drinking water”, and “location”. The first cluster is related to the design of simulation and optimisation

4. Limitations

The work carried out is not free of limitations. Indeed, we used information from a single database with a single search string, limiting the articles we accessed. This may not be enough for further analysis.

5. Conclusion and future research avenues

Within this paper, we carried out a bibliometric review of the research trends of the last ten years in applying Machine Learning to water distribution networks. This analysis has been performed on data retrieved from the WoS database. We obtained significant results, such as the most relevant authors, sources, and documents, and highlighted the collaboration between countries. More than 71% of the publications retained using the PRISMA method were articles with the highest annual growing production spanning the last three years (2020-2022). From the most relevant sources, five stood out for their work: “Journal of Water Resources Planning and Management”, “Water”, “Water Resources Management”, “Water Research”, and “Journal of Hydroinformatics”. This study showed that the most productive countries on the subject with strong collaboration are China and the United States, which are among the developed countries. Only one African country, South Africa, is part of the Scientific Production Countries in 14th place and has some partnerships with India and Ukraine.

Moreover, it appears that the latest applications of Machine Learning are related to water distribution systems design and application of different simulation models. This work highlights two ways for future research: assessing water quality and designing other optimisation models. The application of this research in African countries would be fascinating for a better quality of service and efficient management of this resource, which is inaccessible to many African countries. It would also be interesting to broaden the data used in this work by using several databases and a more extensive year range.

Acknowledgements.

The first author thanks the Regional Scholarship and Innovation Fund (RSIF) for its countless supports.

References

- [1] “The distribution of water on, in, and above the Earth | U.S. Geological Survey.” Accessed: Jun. 03, 2023. [Online]. Available: <https://www.usgs.gov/media/images/distribution-water-and-above-earth>
- [2] “Fiche de résultats : Gestion des ressources en eau,” World Bank. Accessed: May 27, 2023. [Online]. Available: <https://www.banquemonde.org/fr/results/2013/04/15/water-resources-management-results-profile>
- [3] “Les technologies au service d’une gestion intelligente de l’eau - Hello Future Orange,” Hello Future. Accessed: May 27, 2023. [Online]. Available: <https://hellofuture.orange.com/fr/la-gestion-des-reseaux-deau-potable-gagne-en-intelligence/>
- [4] J. Meseguer, J. M. Mirats-Tur, G. Cembrano, and V. Puig, “A decision support system for on-line leakage localization,” *Environ. Model. Softw. Environ. Data News*, vol. 60, pp. 331–345, Oct. 2014, doi: 10.1016/j.envsoft.2014.06.025.
- [5] J. Kang, Y. J. Park, J. Lee, S. H. Wang, and D. S. Eom, “Novel leakage detection by ensemble CNN-SVM and graph-based localization in water distribution systems,” *IEEE Trans. Ind. Electron.*, vol. 65, no. 5, pp. 4279–4289, May 2018, doi: 10.1109/TIE.2017.2764861.
- [6] A. Soldevila, R. M. Fernandez-Canti, J. Blesa, S. Tornil-Sin, and V. Puig, “Leak localization in water distribution networks using Bayesian classifiers,” *J. Process Control*, vol. 55, pp. 1–9, Jul. 2017, doi: 10.1016/j.jprocont.2017.03.015.
- [7] A. Candelieri, “Clustering and Support Vector Regression for Water Demand Forecasting and Anomaly Detection,” *Water*, vol. 9, no. 3, Art. no. 3, Mar. 2017, doi: 10.3390/w9030224.
- [8] G. He, T. Zhang, F. Zheng, and Q. Zhang, “An efficient multi-objective optimization method for water quality sensor placement within water distribution systems considering contamination probability variations,” *Water Res.*, vol. 143, pp. 165–175, Oct. 2018, doi: 10.1016/j.watres.2018.06.041.
- [9] C. de Winter, V. R. Palleti, D. Worm, and R. Kooij, “Optimal placement of imperfect water quality sensors in water distribution networks,” *Comput. Amp Chem. Eng.*, vol. 121, pp. 1–10, 2019, doi: 10.1016/j.compchemeng.2018.10.021.
- [10] R. Guidotti, H. Chmielewski, V. Unnithan, P. Gardoni, T. Mcallister, and J. Lindt, “Modeling the resilience of critical infrastructure: The role of network dependencies,” *Sustain. Resilient Infrastruct.*, vol. 1, pp. 153–168, Nov. 2016, doi: 10.1080/23789689.2016.1254999.
- [11] Q. Xu, R. Liu, Q. Chen, and R. Li, “Review on water leakage control in distribution networks and the associated environmental benefits,” *J. Environ. Sci.*, vol. 26, no. 5, pp. 955–961, May 2014, doi: 10.1016/S1001-0742(13)60569-0.
- [12] S. Rathi, R. Gupta, and L. Ormsbee, “A review of sensor placement objective metrics for contamination detection in water distribution networks,” *Water Supply*, vol. 15, no. 5, pp. 898–917, Jun. 2015, doi: 10.2166/ws.2015.077.
- [13] A. Garzón, Z. Kapelan, J. Langeveld, and R. Taormina, “Machine Learning-Based Surrogate Modeling for Urban Water Networks: Review and Future Research Directions,” *Water Resour. Res.*, vol. 58, no. 5, p. e2021WR031808, 2022, doi: 10.1029/2021WR031808.
- [14] T. Matthews, “LibGuides: Web of Science platform: Introduction.” Accessed: May 22, 2023. [Online]. Available: <https://clarivate.libguides.com/webofscienceplatform/introduction>
- [15] A. Mukherjee, S. S. Babu, and S. Ghosh, “Thinking about water and air to attain Sustainable Development Goals during times of COVID-19 Pandemic,” *J. Earth Syst.*

- Sci.*, vol. 129, no. 1, p. 180, Aug. 2020, doi: 10.1007/s12040-020-01475-0.
- [16] “Aims & Scope and Editorial Board: Journal of Water Resources Planning and Management | ASCE Library,” Journal of Water Resources Planning and Management. Accessed: May 22, 2023. [Online]. Available: <https://ascelibrary.org/page/jwrmd5/editorialboard?doi=10.1061%2Fjwrmd5&publicationCode=jwrmd5>
- [17] “Water.” Accessed: May 22, 2023. [Online]. Available: <https://www.mdpi.com/journal/water/about>
- [18] “Avi Ostfeld, Technion | Updated: April 13, 2023.” Accessed: May 22, 2023. [Online]. Available: <https://ostfeld.net.technion.ac.il/>
- [19] I. V. Tetko, V. Kůrková, P. Karpov, and F. Theis, Eds., *Artificial Neural Networks and Machine Learning – ICANN 2019: Text and Time Series: 28th International Conference on Artificial Neural Networks, Munich, Germany, September 17–19, 2019, Proceedings, Part IV*, vol. 11730. in Lecture Notes in Computer Science, vol. 11730. Cham: Springer International Publishing, 2019. doi: 10.1007/978-3-030-30490-4.
- [20] T. P. Lambrou, C. C. Anastasiou, C. G. Panayiotou, and M. M. Polycarpou, “A Low-Cost Sensor Network for Real-Time Monitoring and Contamination Detection in Drinking Water Distribution Systems,” *IEEE Sens. J.*, vol. 14, no. 8, pp. 2765–2772, Aug. 2014, doi: 10.1109/JSEN.2014.2316414.
- [21] “Evaluating risk of water mains failure using a Bayesian belief network model - ScienceDirect.” Accessed: Nov. 15, 2023. [Online]. Available: <https://www.sciencedirect.com/science/article/abs/pii/S0377221714005360>
- [22] N. Quinete and R. A. Hauser-Davis, “Drinking water pollutants may affect the immune system: concerns regarding COVID-19 health effects,” *Environ. Sci. Pollut. Res.*, vol. 28, no. 1, pp. 1235–1246, Jan. 2021, doi: 10.1007/s11356-020-11487-4.
- [23] “CDC Museum COVID-19 Timeline,” Centers for Disease Control and Prevention. Accessed: May 23, 2023. [Online]. Available: <https://www.cdc.gov/museum/timeline/covid19.html>
- [24] L. A. Rossman, *Epanet 2: Users Manual*, vol. 38. Cincinnati: US Environmental Protection Agency National Risk Management Research Laboratory, 2000. [Online]. Available: <https://books.google.bj/books?id=CgijMwEACAAJ>
- [25] A. Ostfeld, M. Asce, and A. Tubaltzev, “Ant Colony Optimization for Least-Cost Design and Operation of Pumping Water Distribution Systems,” *J. Water Resour. Plan. Manag.-Asce - J WATER RESOUR PLAN MAN-ASCE*, vol. 134, Mar. 2008, doi: 10.1061/(ASCE)0733-9496(2008)134:2(107).
- [26] R. Puust, Z. Kapelan, D. A. Savic, and T. Koppel, “A review of methods for leakage management in pipe networks,” *Urban Water J.*, vol. 7, no. 1, pp. 25–45, Feb. 2010, doi: 10.1080/15730621003610878.