

NRC Publications Archive Archives des publications du CNRC

Scientometric insights into the trends and evolution of the underwater sensing technologies and their applications

Ebadi, Ashkan; Auger, Alain; Gauthier, Yvan

This publication could be one of several versions: author's original, accepted manuscript or the publisher's version. / La version de cette publication peut être l'une des suivantes : la version prépublication de l'auteur, la version acceptée du manuscrit ou la version de l'éditeur.

For the publisher's version, please access the DOI link below. / Pour consulter la version de l'éditeur, utilisez le lien DOI ci-dessous.

Publisher's version / Version de l'éditeur:

<https://doi.org/10.5530/jscires.13.2.44>

Journal of Scientometric Research, 13, 2, pp. 562-574, 2024-08-19

NRC Publications Archive Record / Notice des Archives des publications du CNRC :

<https://nrc-publications.canada.ca/eng/view/object/?id=ca53a183-cd99-45be-a5f2-2f63cc963752>

<https://publications-cnrc.canada.ca/fra/voir/objet/?id=ca53a183-cd99-45be-a5f2-2f63cc963752>

Access and use of this website and the material on it are subject to the Terms and Conditions set forth at

<https://nrc-publications.canada.ca/eng/copyright>

READ THESE TERMS AND CONDITIONS CAREFULLY BEFORE USING THIS WEBSITE.

L'accès à ce site Web et l'utilisation de son contenu sont assujettis aux conditions présentées dans le site

<https://publications-cnrc.canada.ca/fra/droits>

LISEZ CES CONDITIONS ATTENTIVEMENT AVANT D'UTILISER CE SITE WEB.

Questions? Contact the NRC Publications Archive team at

PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca. If you wish to email the authors directly, please see the first page of the publication for their contact information.

Vous avez des questions? Nous pouvons vous aider. Pour communiquer directement avec un auteur, consultez la première page de la revue dans laquelle son article a été publié afin de trouver ses coordonnées. Si vous n'arrivez pas à les repérer, communiquez avec nous à PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca.

Scientometric Insights into the Trends and Evolution of the Underwater Sensing Technologies and their Applications

Ashkan Ebadi^{1,*}, Alain Auger², Yvan Gauthier¹

¹Digital Technologies, National Research Council-Canada, CANADA.

²Science and Technology Foresight and Risk Assessment Unit, Defence Research and Development-Canada, CANADA.

ABSTRACT

Underwater sensing has a wide range of applications including the location and monitoring of subsurface infrastructures such as cables and pipelines, mapping undersea terrain, the study of marine life, pollution and salinity monitoring and detection of seismic activities. Underwater sensing technologies are also essential to maritime surveillance, detection and tracking of submarines and other underwater objects. Using natural language processing and machine learning, we mined scientific publications to gain insight into underwater sensing research evolution during the 21st century, in terms of technology development and applications. This study offers a comprehensive analysis of the underwater sensing research landscape as well as the temporal evolution of its main research topics. We identified 18 key research topics that offer comprehensive and logical coverage of the research field. Nearly half of them are on a decreasing trend, despite an overall increase in the number of scientific publications. These findings and the extracted patterns can provide researchers and decision-makers with new insights into the field, its characteristics and its development. Our proposed algorithmic approach can also be applied to other areas for technologies of a disruptive nature.

Keywords: Technology trends, Research topic evolution, Underwater sensing, Machine learning, Topic modeling.

Correspondence:

Ashkan Ebadi

National Research Council-Canada,
CANADA.

Email: ashkan.ebadi@nrc-cnrc.gc.ca

ORCID: 0000-0002-4542-9105

Received: 23-10-2023;

Revised: 05-03-2024;

Accepted: 28-03-2024.

INTRODUCTION

The surface of our planet is mostly covered by the ocean, occupying 71% of the total surface area.^[1] Oceans' ecosystems are greatly linked to the land which makes studying and understanding the oceans extremely important.^[2] This can, for example, help to better manage environmental protection and climate change issues. In addition to the ecological connectivity between the sea and the land, oceans are home to rich resources, such as minerals, hydrocarbons and biological resources.^[3,4] For that reason, underwater sensing technologies have attracted huge interest from not only the scientific community but also various private and public-sector organizations. These technologies also have several defence and security applications, such as maritime surveillance,^[5,6] mine detection,^[7] communication cable protection,^[8] or anti-submarine warfare.^[9]

Ocean exploration is traced back to the 19th century when the HMS Challenger made the first oceanographic expedition,^[10] but from that time until now, the technology and exploration methods

have drastically advanced. Underwater sensing technologies based on various sensing methods such as acoustics, optics and electromagnetics, are now being widely applied for ocean exploration.^[11] Acoustic sensing employs sonar devices for various purposes such as submersible navigation and seafloor mapping. Optical sensing technologies include but are not limited to underwater object detection and inspection, spectrophotometry and fluorophotometry. Electromagnetics technologies are, for instance, applied for the inspection and detection of underwater cables and pipelines.^[2]

Each of these technologies has its own advantages and drawbacks. For instance, underwater acoustic sensors have a high range but a low resolution, whereas optical sensors can achieve high-resolution images but at a short range underwater.^[11] That is why most underwater sensing applications involve systems of networked sensors, but such networks face challenges around bandwidth, propagation delays, routing, power constraints and others.^[12]

Given rapid developments in underwater sensing technologies, monitoring their evolution is essential to better understand how they could impact economic development,^[13] environmental protection^[2] and national security.^[6,7,9] To this end, we use an advanced probabilistic topic modeling method and combine natural language processing and machine learning techniques



DOI: 10.5530/jscires.13.2.44

Copyright Information :

Copyright Author (s) 2024 Distributed under
Creative Commons CC-BY 4.0

Publishing Partner : EManuscript Tech. [www.emanuscript.in]

to extract the main research themes in the field of underwater sensing and statistically examine their temporal changes. By mining thousands of scientific publications, we offer a comprehensive analytic framework able to reveal the target research landscape as well as the temporal evolution of its main research topics. Our results uncover patterns that can provide researchers and decision-makers with new insights into the field, its characteristics, areas of focus and its development. The main research questions in this study are as follows:

What is the main research themes applied to underwater sensing that the scientific community has paid attention to in the 21st century?

What research themes have increased/decreased statistically?

Which research themes are likely to show increased research activity in the future?

The remainder of this paper is as follows. Section “Data and methodology” describes data and techniques in detail. Section “Results” presents the findings of the research. Findings are then discussed and the conclusions are presented in Section “Discussion and conclusion”. Finally, some limitations of the research and future directions are presented in Section “Limitations and future work”.

Data and methodology

Data

In this study, we selected Elsevier's Scopus as the source of data. Using a comprehensive search query (refer to Appendix A.), carefully designed by our domain experts, we collected all scientific publications about *underwater sensing technologies* ($n=10,852$). Data were collected on September 14, 2022, hence; data of 2022 do not cover the whole year. The collected data contained all the meta-data about each of the extracted publications that were available on the Scopus website. This included but was not limited to the titles and abstracts of the papers date of publication, authors and authors' affiliations. We only included papers that were written in English. We removed rows with no publication date available ($n=53$). Publications with no title or abstract were excluded ($n=78$) and duplicate records ($n=173$) were removed, resulting in a dataset containing 10,548 records. And, finally we filtered in papers published within the period from 2000 to 2022 ($n=10,113$).

After collecting the publications data, we first generated a new feature, called “*pubtext*”, by combining the title and the abstract for each publication in the dataset. The abstract summarizes the full contribution or content of a publication in a short and definitive manner, hence; it contains a lot of useful information. Despite being very short and limited, the title of a publication may also provide some complementary informative keywords and/or keyphrases that may not necessarily be present in the

abstract section. These reasons motivated us to merge titles and abstracts to obtain a more precise representation of the publications' content. Scopus does not provide the full text of publications. We applied several preprocessing steps to *pubtext* as follows: 1) lowercase conversion, 2) stop words removal, using a comprehensive custom list that included common English stop words plus a list of curated lists of stop words generated by our team and tailored for publication data analysis ($n=1,983$), 3) special character correction, 4) punctuation removal, 5) tokenization and 6) stemming, i.e., reducing terms (tokens) to their root form. Using the processed text data, we then created a document-term frequency matrix in which rows represent publications, columns represent the tokenized stems and each cell value is the number of appearances of a given term in a given publication. The final processed dataset comprises 10,113 publication records, encompassing 4,359 articles (including those in press and data papers), 5,199 conference papers and 210 books and book chapters. Additionally, the dataset includes 345 records categorized as reviews, surveys, letters and editorials. The methodological framework is explained in the next section in detail.

METHODOLOGY

The analytics engine of this study contains three main components, i.e., temporal trend verification, temporal text complexity analysis and topic evolution analysis. In this section, these components are explained in detail.

Temporal trend verification

Prior to analyzing the temporal evolution of research topics in scientific publications about underwater sensing technologies, we first verified the existence of the temporal trend. For this purpose, we applied the correspondence analysis (CA)^[14] to the generated Document-Term Frequency (DTF) matrix, as explained in Section “Data”. The generated DTF matrix was first split into 23, i.e., the number of the annual intervals in [2000,2022], Term-Frequency (TF) matrices by filtering on the year of publications. To reduce noise and increase the precision, we excluded terms for which frequency was less than 20 each year. To improve specificity, we also removed common terms, i.e., terms that were present in more than 60% of the publications each year. The filtered TF matrices were combined and using the CA algorithm, were mapped to a 2-dimension (2D) space by extracting the first 2 principal components. The 2D map provides a visually understandable representation in which noises are filtered out and could help to understand general temporal patterns in the data and verify the existence of a temporal trend.

Textual similarity analysis

After verifying the existence of temporal evolution in the underwater sensing literature, the textual similarity between publications in different years was investigated. This descriptive

analysis helps to evaluate the evolution of research terminology in the examined field. For this purpose, we took two approaches as follows: 1) a common bag of words approach using the Term Frequency-Inverse Document Frequency (TF-IDF) vectorizer and 2) a modern contextual approach, i.e., Cross Encoder.^[15]

In the bag of words approach, TF-IDF represents each document in a corpus as a vector of word frequencies and it calculates the importance of each word in a document relative to the entire corpus. In other words, TF-IDF measures the frequency of a word within a document (TF) and the rarity of the word across the entire corpus (IDF), without considering the context in which each word appears. In this approach and by using TF-IDF, documents with similar word distributions are considered similar, even if the order and context of words may vary. In addition to not being sensitive to the context, the vectors generated by the bag of words approaches are high-dimensional and sparse, especially if the number of documents is large.

Due to the mentioned drawbacks of the bag of words approach and to complement our findings, we also used a modern contextual approach, i.e., Cross Encoder that considers the context in which different words appear to measure textual similarity. We used Hugging Face *stsb-roberta-base* model^[15] which is a pre-trained sentence transformer model based on the RoBERTa architecture,^[16] short for “Robustly optimized BERT approach” and is fine-tuned for the Semantic Textual Similarity (STS) task. Unlike TF-IDF, which focuses on word frequencies, the CrossEncoder looks at the overall meaning of the texts, making it more suitable for capturing nuanced similarities and contextual understanding between documents in the corpus.

Topic extraction and evolution analysis

As the last step in analyzing the temporal evolution of *underwater sensing* publications, we applied the *Structural Topic Modeling* (STM) technique^[17] to extract latent research topics and estimate their temporal evolution. Topic Modeling (TM) is an unsupervised machine learning technique that can automatically summarize massive textual data by extracting latent semantic themes, capturing the most prevalent subjects.^[18] TM has been widely used across many disciplines and in many applications such as economics,^[19] information sciences,^[20] social media analytics,^[21] cancer research,^[22] bio-informatics^[23] and new diseases/pandemics,^[24] to name a few.

Unlike static TMs that capture the topics at one moment, STM is a temporal TM that allows capturing the evolution of topics across time by incorporating document-level covariates of interest. In our case, we incorporated the date of publications into the model as the covariate of interest to analyze the temporal evolution of the *underwater sensing* domain. For this purpose, the proportion of each publication was regressed on the date of publication. STM also calculates correlations between the extracted topics which

help to better realize the dynamics of the examined domain. In addition, similar to other TM techniques, STM requires no data labeling. Motivated by these properties, we built an STM model with an annual granularity on underwater sensing scientific publications.

One of the main parameters of the algorithm is the number of topics that should be set prior to the model-building process. Finding the optimal number of topics quantitatively and automatically is a challenging task,^[25] often leading to inaccurate results.^[26,27] To overcome the limitations of a fully automatic/manual approach, we applied a semi-automatic approach to find the optimal number of topics. First, we built several baseline topic models, varying the number of topics in the range.^[5,30] The top keywords/keyphrases associated with each topic of the baseline models as well as topic-word distributions were then analyzed. Intuitively, topics of higher quality are more interpretable and have fewer overlapping keywords.^[28] To further narrow down the range for the number of topics, we used an intrinsic topic coherence evaluation metric, i.e., the C_v score.^[29] We found the optimal number of topics to be in the range of.^[16,20] We then built 5 topic models with topic numbers in the mentioned range, verified their keywords and concluded that the optimal number of topics for the examined corpus is 18.

STM does not generate a label for the extracted topics which makes the interpretation of the topics difficult, especially for non-domain experts. We employed a multi-layer approach to automatically generate representative labels for the extracted topics. The approach had 5 steps as follows: 1) a random set of papers were selected for each of the extracted topics, 2) the selected papers were summarized using a pre-trained Text-To-Text Transfer Transformer (T5) model,^[30] 3) the summaries were combined, 4) using a pre-trained T5-based headline generator model, a representative title was generated for the combined summaries, 5) we repeated steps 1 to 4 fifteen times to generate fifteen representative labels per topic. To ensure that the extracted topics represent the examined field properly, the extracted topics as well as the automatically generated labels were reviewed and validated by senior scientists with expertise in underwater sensing research from Defence Research and Development Canada and a single topic label was assigned to each topic. Figure 1 shows the conceptual flow of the analysis.

The pipeline contains 3 main components: 1) data collection that collects underwater sensing publications within the period of [2000, 2022], 2) data engineering and processing that pre-processes the collected data and makes it ready for analysis and 3) data analytics that first verifies the existence of a temporal trend. It then performs a textual similarity analysis using two different approaches. And finally, a structural topic model is built, research topics are extracted and automatically labelled and the evolution of the extracted research topics as well as their correlation is analyzed. The automatically generated labels and

topics' quality are verified by domain experts to ensure their validity.

RESULTS

Descriptive analysis

By analyzing the authors' affiliations, the top countries that have been working on underwater sensing in the examined data set were extracted. The top-3 countries are: 1) China ($n = 3190$), 2) the USA ($n = 1956$) and 3) the UK ($n = 838$). Canada stands in

the 4th place with 431 papers. Figure 2 shows the distribution of underwater sensing publications, with navy blue and red bars representing the total number of publications and Canadian publications, respectively. As seen in the figure, although the general trend is increasing, publications on underwater sensing demonstrate a fluctuation, possibly attributable to the rapidly evolving and interdisciplinary nature of underwater sensing technologies. At the beginning of the examined period, from 2000 to 2003, the number of publications was few. However, the

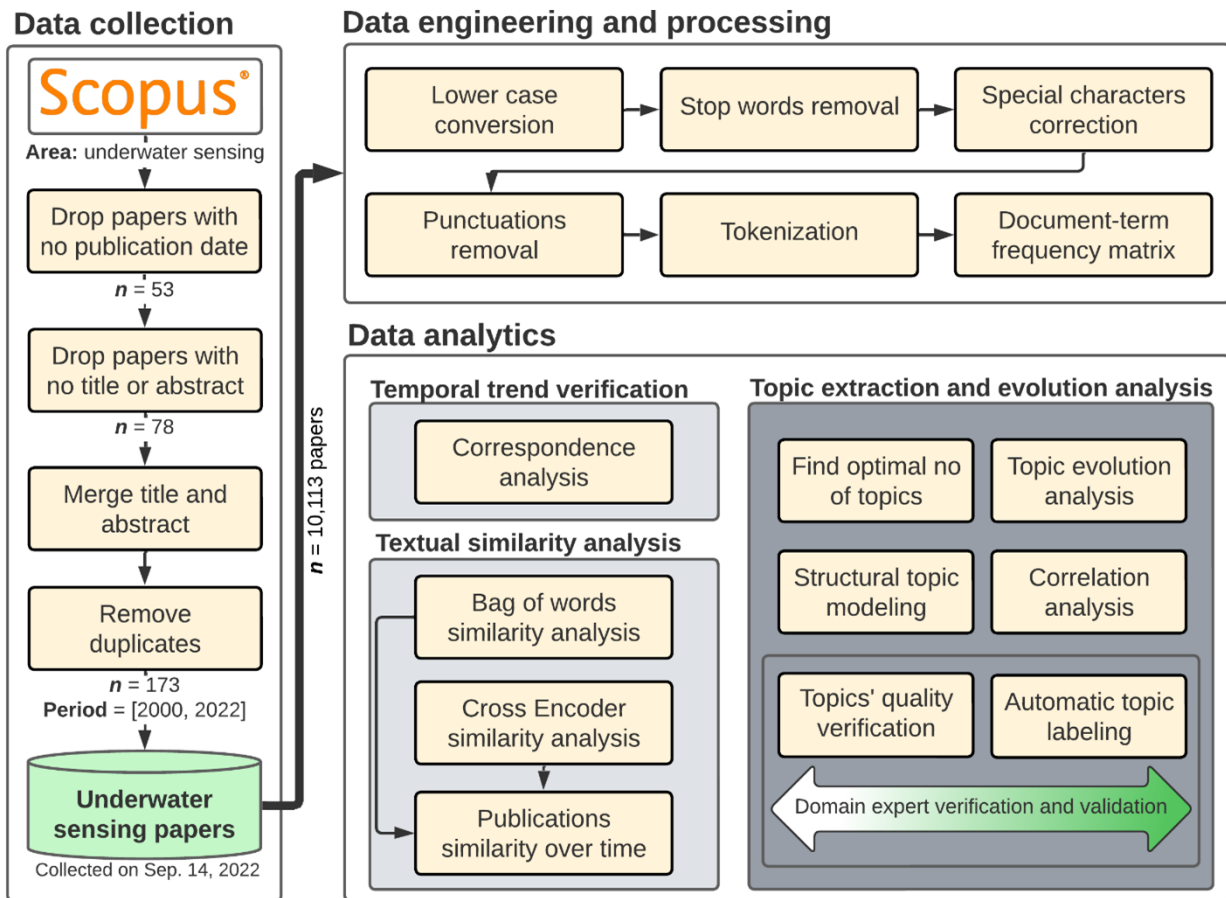


Figure 1: The high-level conceptual flow.

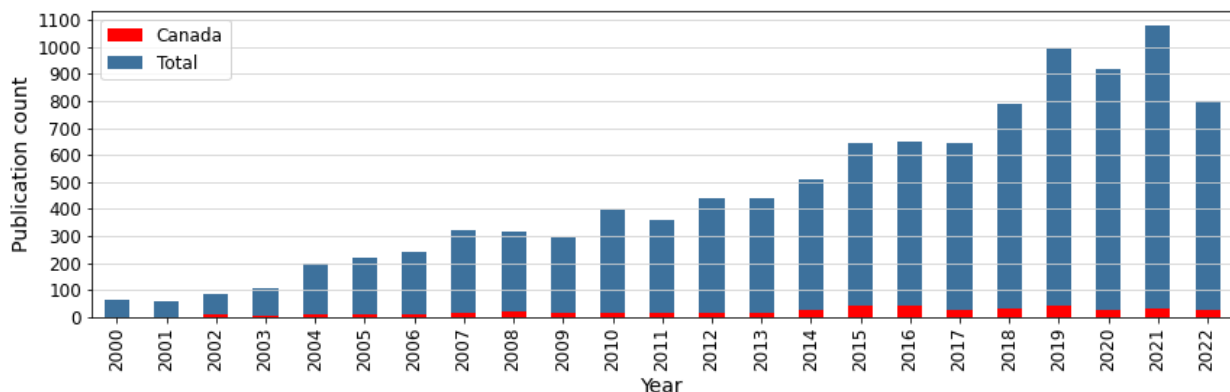


Figure 2: The trend of underwater sensing publications between 2000 and 2022.

number of publications has increased over time. The increase rate was steeper in recent years, with a more than double increase from 2014 to 2021. The increasing rate of publications is a common trend and has been observed in other domains as well (e.g.,^[27,31]). It should be noted that the observed growth could be partially due to higher coverage of publications in recent years.

We did an initial exploratory analysis on the dynamics of the examined field by performing a co-word analysis.^[32] Co-word analysis may help to extract patterns and identify trends.^[33] Figure 3 shows the co-occurrence networks of top-20 bigrams for year in [2000, 2005, 2010, 2015, 2020 and 2022]. To refine the figures further, we filtered out bigrams in which both terms were among

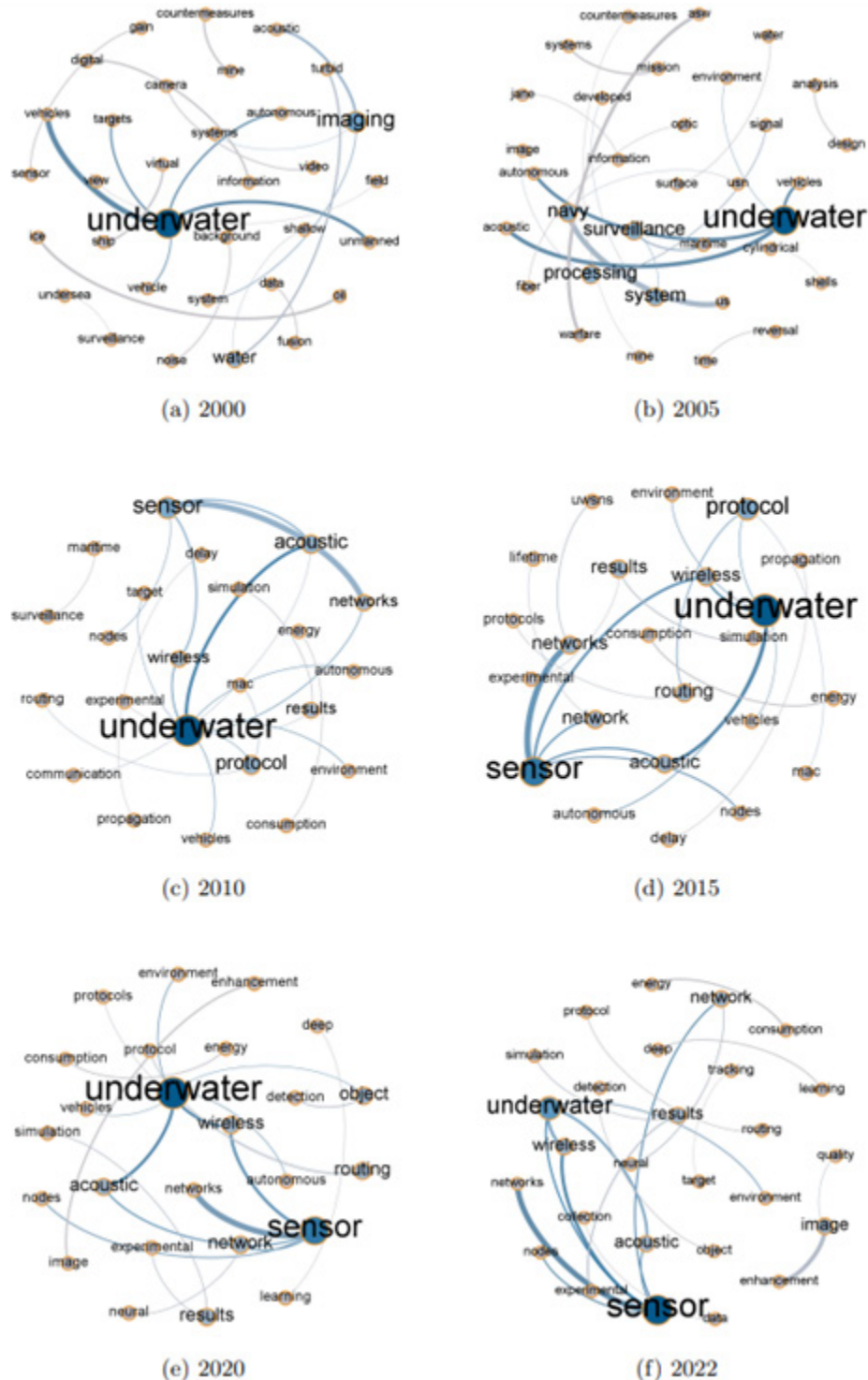


Figure 3: Co-occurrence networks. Node size represents the degree of the node, i.e., the number of direct connections. Font size is proportionate to the degree of the node.

the top-3 most frequent terms, i.e., [*underwater, sensor*, imag**]. As seen, *underwater* has been the term with the highest degree, i.e., has been present in the most bigrams, in all examined periods except for 2022 when *sensor* took its place.

Temporal trend verification

As the first step and before performing a temporal evolution analysis, we verified the existence of a temporal trend in the collected underwater sensing publications by running a correspondence analysis on the most frequent terms extracted from the publications. Figure 4 shows the result where a U-shape curve is observed for years in the examined period. Based on the distance of years, indicated by orange points in the figure, from the axes and origin, the existence of a temporal trend is confirmed. In addition, it is seen that papers published in 2004 and 2005 are the farthest from the origin which may indicate a relatively different research terminology/focus in the mentioned years.

Textual similarity analysis

As explained in Section “Textual similarity analysis”, a conventional TF-IDF-based bag of words approach and an advanced transformer-based approach, i.e., Cross Encoder, were used to analyze publications' similarity over time. Figure 5-a shows the results of the textual similarity analysis using a bag of words TF-IDF vectorizer. As seen, in general, similarities between publications increased over time with publications in recent years

reflecting the highest similarity. Based on the performed analysis, this increasing similarity shift seems to emerge around 2010. Publications in 2002-2009 have had the lowest similarity to their previous publications. And, research published in 2004-2005 has been less used in the following years. This is in line with our findings in the previous section where a different terminology/research focus was observed for research published in 2004-2005. This could be an indication of a shift in research priorities.

To further investigate and provide a different perspective, we also did textual similarity analysis using a modern contextual approach, i.e., Cross Encoder (as explained in Section “Textual similarity analysis”). Results are depicted in Figure 5-b. As seen, research published in 2016 has been the most similar to studies in previous years, i.e., 2000-2015, while the similarity decreased after 2016. Similarity scores of publications in 2022 are also noticeable. In this respect, publications in 2020 and 2021 are of note as they have been the least similar to previous periods. Overall, a sinusoidal pattern of similarity is observed where a relatively high similarity in a couple of years is followed by a relatively low similarity in the following period.

Structural topic modeling

Topic modeling results are presented in this section. First, the extracted topics are listed. Next, the topics' correlations are analyzed. And, finally, the temporal evolution of the main research themes in the field of underwater sensing is investigated.

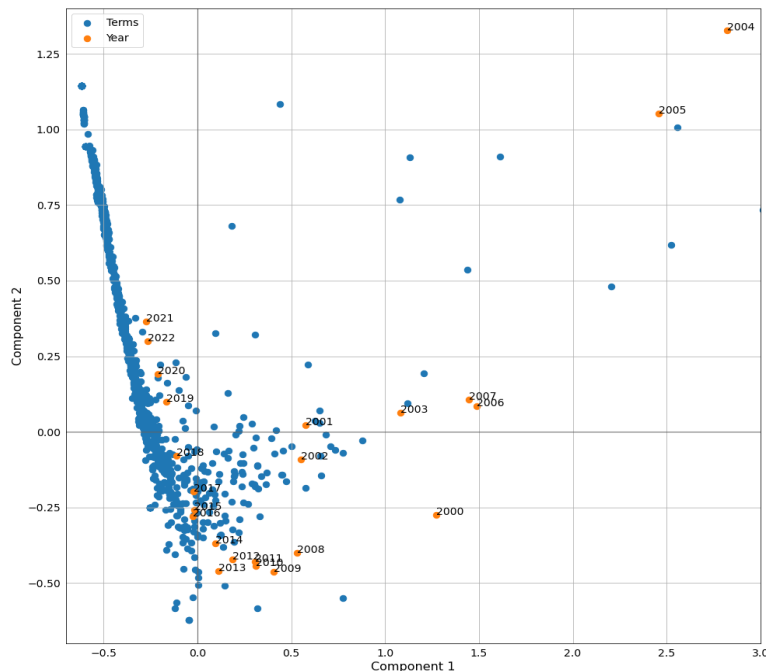


Figure 4: Temporal trend verification by running a correspondence analysis on the frequent terms that appeared in the underwater sensing publications from 2000 to 2022. Orange points represent years. Blue points represent the frequent terms.

Extracted topics

As explained in Section “Topic extraction and evolution analysis”, we used structural topic modeling^[17] to extract 18 research themes from the collected underwater sensing publications. The year of publication was used as the covariate which was used to assess the temporal evolution of publications. The extracted topics are as follows:

- Topic 1:** Underwater imaging
- Topic 2:** Wireless sensor networks
- Topic 3:** Autonomous Underwater Vehicles (AUVs)
- Topic 4:** Underwater image processing techniques
- Topic 5:** Sonar for target/submarine detection
- Topic 6:** Sensor systems
- Topic 7:** Maritime security
- Topic 8:** Machine/deep learning for underwater object detection/classification
- Topic 9:** Network architectures, protocols and routing
- Topic 10:** Maritime anomaly detection
- Topic 11:** Underwater optical imaging
- Topic 12:** Underwater sensor networks
- Topic 13:** Underwater acoustic communications
- Topic 14:** Target tracking
- Topic 15:** Subsea sensing and monitoring
- Topic 16:** Anti-submarine warfare
- Topic 17:** Marine environment

Topic 18: Underwater research and applications

Figure 6 shows the word clouds of the above-listed topics using stems of their representative keywords. The stem of a term is basically its root form, which helps group together different variations of the same word. For example, if we have the words “detecting”, “detected”, and “detects”, their stem would be “detect”. We set the maximum number of words to be plotted at 120 and set the minimum frequency for a term at 3. The size of the term represents its importance in the given topic that is the larger the term the more common it was.

Topics correlation analysis

By showing interrelationships between the extracted topics, Figure 7 provides insights on topics that are likely to co-appear in a set of publications. As seen in Figure 7-a, most of the extracted topics are negatively correlated, e.g., (Topic 1, Topic 2), meaning that it is unlikely for them to occur in the same publication set. We can also observe pairs with no or very minor correlation, e.g., (Topic 3, Topic 15). Figure 7-b depicts the correlation graph of the extracted topics, showing only positive correlations. Node size represents the coverage of the topic and the thickness of the edge indicates the correlation coefficient. The highest positive correlation is observed for (Topic 2, Topic 9), followed by (Topic 2, Topic 12), (Topic 2, Topic 18) and (Topic 1, Topic 4) pairs in descending order. These interrelationships are quite expected. Topic 2 is about wireless sensor networks which are related to network architectures and protocols (Topic 9), underwater sensor networks (Topic 12) and underwater research and applications (Topic 18). These highly correlated topics are well illustrated by the increasing research trend towards the implementation of the Internet of Underwater Things (IoUT).^[34,35] Underwater imaging (Topic 1) is also positively correlated to image processing techniques (Topic 4).

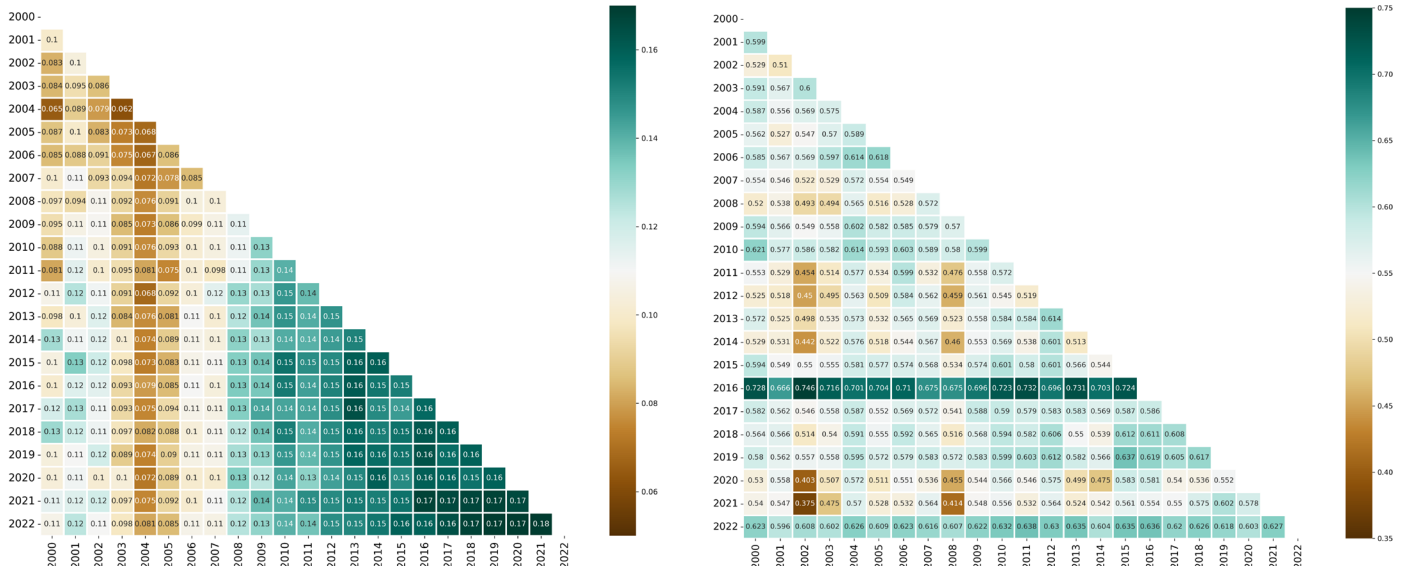


Figure 5: Textual similarity analysis.



Figure 6: Word clouds of the extracted topics. Font size represents the importance of the term. The numbers in parentheses located in the top left corner of each word cloud indicate the corresponding topic number.

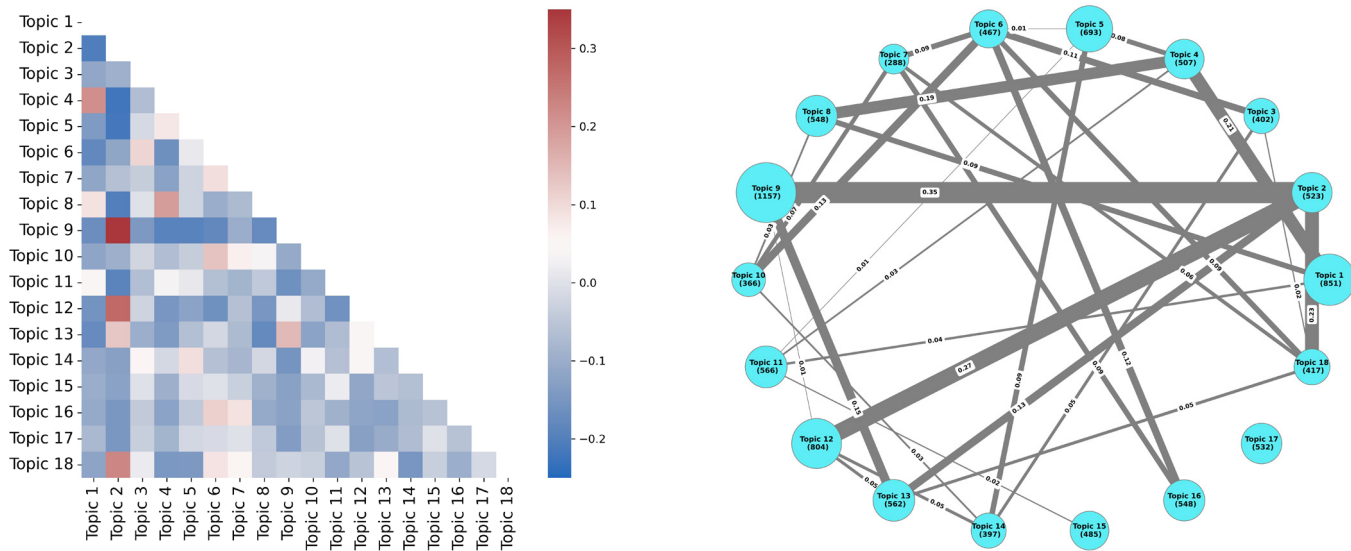


Figure 7: Topics correlations. Node size in the graph represents the coverage of a topic. The thickness of the edges in the graph represents the correlation coefficient.

Temporal evolution analysis

Figure 8 shows the expected prevalence of the extracted topics over time. Shaded areas in the figure indicate 95% confidence intervals. As observed, 10 topics (i.e., Topics 1-2, 4, 7-10, 12, 14-15 and 18) followed an increasing trend during the examined period. The increasing slope for Topic 15, i.e., subsea sensing and monitoring, is lower than others. Underwater imaging (Topic 1) and research on image processing techniques (Topic 4) experienced a sharp increase. This combined with the increase observed in employing AI techniques for underwater object detection/classification (Topic 8) and underwater research and applications (Topic 18) may indicate a recent research direction toward applying AI and computer vision to the field of underwater sensing. Wireless and underwater sensor networks (Topics 2 and 12) also experienced an increasing trend. Maritime anomaly detection (Topic 10) also increasingly attracted the attention of researchers. The other 8 topics (i.e., Topics 3, 5-7, 11, 13 and 16-17) experienced a decreasing trend. Despite the increase observed in advanced image processing and computer vision techniques, research on underwater optical imaging (Topic 11) followed a decreasing trend. The decrease in marine environment research (Topic 17) and AUVs (Topic 3) are noticeable.

DISCUSSION AND CONCLUSION

Understanding and analyzing the landscape of underwater sensing technologies is of paramount importance due to the critical role these technologies play in addressing various environmental, scientific and industrial challenges. Underwater sensing facilitates the monitoring and conservation of marine ecosystems, enables efficient resource exploration, aids in disaster response and recovery efforts and enhances our understanding of climate change impacts on oceans. Moreover, these technologies also have several defence and security applications, such as

maritime surveillance, mine detection, communication cable protection and anti-submarine warfare.

The field of underwater sensing research is a large one and evolving that would be difficult to properly characterize without an automated and systematic approach. Our proposed AI-powered solution identified 18 key topics that offer a comprehensive and logical coverage of the research field. The amount of scientific publications has been growing in almost all disciplines^[27,31] and underwater sensing is no exception. However, our results show that nearly half of the topics identified (8 out of 18) are on a decreasing trend.

Overall, a decrease/increase in a specific research interest is a natural part of the scientific process, as researchers continuously seek new challenges and opportunities for advancements in the field. The observed temporal decrease could be due to several reasons such as saturation of the topic, technological limitations, shifting priorities, emergence of new trends, market dynamics and lack of demands, to name a few. On the other hand, a research topic may experience a temporal increase for various reasons such as emerging challenges and pressing issues, technological advancements, cross-disciplinary applications, policy and regulatory changes, collaboration and knowledge exchange and academic interest, to name a few. Although understanding the reasons behind these declines/inclines can inform decision-makers on how to better foster innovation in underwater sensing technologies, this would be difficult to investigate due to the complex dynamics in research activities.

As the field of underwater sensing is continuously evolving with the emergence of new technologies and research directions, decision-makers face the challenge of keeping up with the rapidly changing landscape. The fact that the analysis workflow is largely automated is significant and the proposed AI-powered

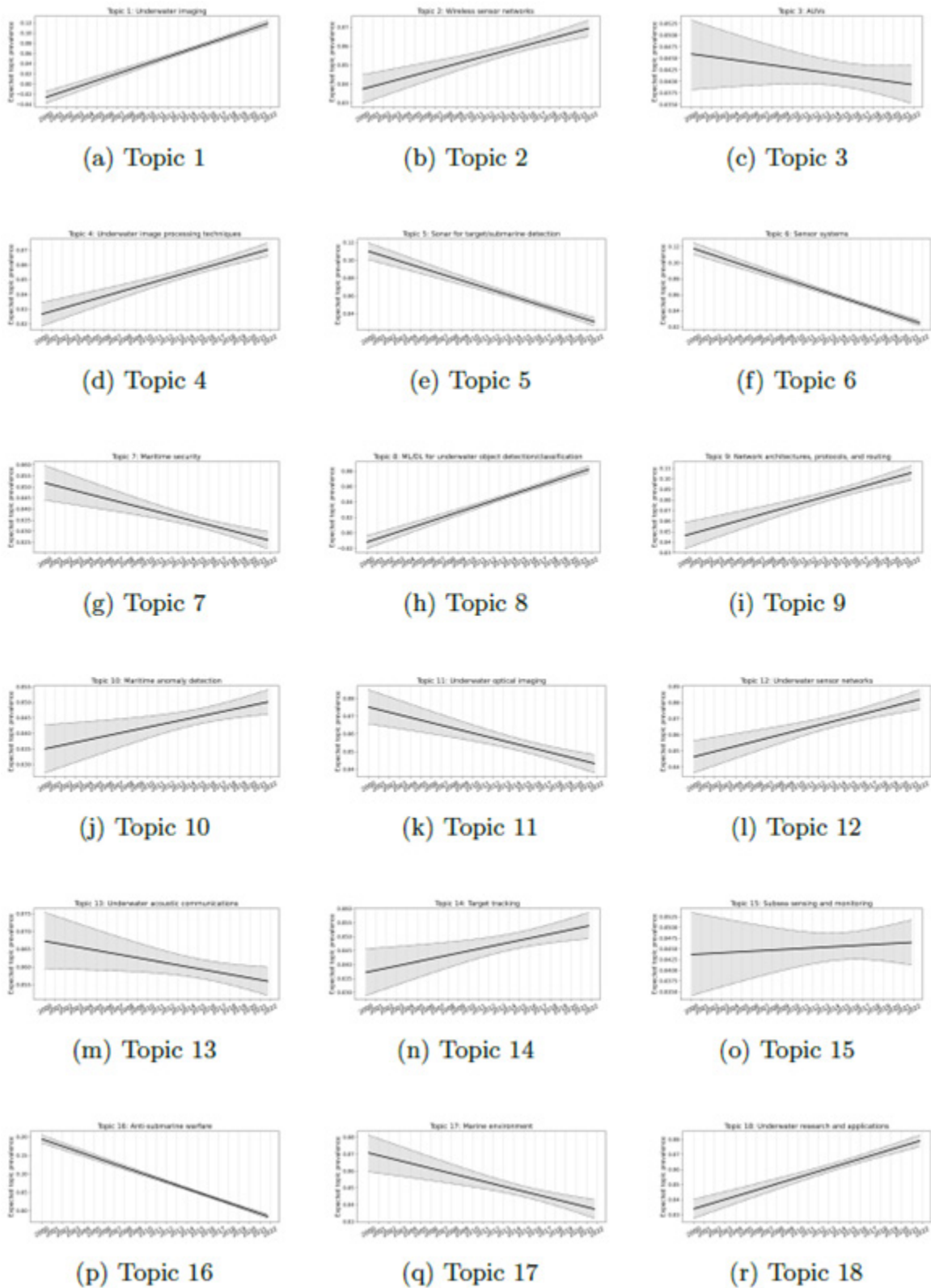


Figure 8: Temporal evolution of main research themes in underwater sensing. The shaded areas between the dotted lines indicate the 95% confidence interval.

approach, leveraging natural language processing and machine learning, can significantly assist decision-makers by providing a comprehensive and up-to-date analysis of the temporal evolution of research topics in underwater sensing. Extracting topics algorithmically allows us to process thousands of publications, while removing subjectivity from the analysis and enabling consistent comparisons between topics and between time intervals. By identifying emerging trends and highlighting the most relevant areas of investigation, this approach can empower decision-makers to make informed choices regarding resource allocation, investment and policy formulation, ultimately leading to more effective and sustainable advancements in underwater sensing technologies.

Although we focused on underwater sensing as the case technology, the methodology utilized in this study shows promise for broader applicability beyond underwater sensing. It has the potential to be employed in various domains marked by disruptive technologies to provide insights into research landscapes recognize emerging trends and guide decision-making processes. Automation also means that this kind of insight can be generated quickly, even for complex research fields that produce thousands of papers every year. Accordingly, it allows analysts to better understand the dynamics of a particular research field and make evidence-based recommendations to decision-makers from the public and private sectors. This is particularly important for technologies such as underwater sensing, given the environmental, economic and military applications of these technologies.

LIMITATIONS AND FUTURE WORK

We used a collection of scientific publications to study the characteristics and landscape of underwater sensing research over the period from 2000 to 2022. Data for 2022 was not complete at the time of our data collection, however, we decided to include it in the study due to multiple reasons such as the high volume of publications and comparison purposes. Future research can use more recent data. Future research may also consider using other data sources, e.g., patents. Our proposed pipeline extracts high-level research themes. Future work may consider other levels of abstraction. Finally, we did not have access to the full text of publications. Analyzing other sections of publications, e.g., the method section, and the full text could be a potential future research direction.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

AVAILABILITY OF DATA AND MATERIAL

Data was collected from Elsevier's Scopus using the search query listed in Appendix A.

CODE AVAILABILITY

Unfortunately, we are not able to share the code due to the intellectual property involved. However, interested readers can contact the corresponding author if they have any questions regarding the pipeline.

REFERENCES

- Wang Y, Liu Y, Guo Z. Three-dimensional ocean sensor networks: A survey. *J Ocean Univ China*. 2012;11(4):436-50.
- Sun K, Cui W, Chen C. Review of Underwater Sensing Technologies and Applications. *Sensors*. 2021;21(23):7849.
- Hwang J, Bose N, Nguyen HD, Williams G. Acoustic Search and Detection of Oil Plumes Using an Autonomous Underwater Vehicle. *J Mar Sci Eng*. 2020;8(8):618.
- Zhai M, Hu R, Wang Y, Jiang S, Wang R, Li J, *et al.* Mineral Resource Science in China: Review and perspective. *Geogr Sustain*. 2021;2(2):107-14.
- Melodia T, Kulhandjian H, Kuo LC, Demirors E. Advances in Underwater Acoustic Networking. In: Basagni S, Conti M, Giordano S, Stojmenovic I, editors. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 2013. p. 804-52. DOI: 10.1002/9781118511305.ch23
- Terracciano DS, Bazzarello L, Caiti A, Costanzi R, Manzari V. Marine Robots for Underwater Surveillance. *Curr Robot Rep*. 2020;1(4):159-67.
- Hozyń S. A Review of Underwater Mine Detection and Classification in Sonar Imagery. *Electronics*. 2021;10(23):2943.
- Eleftherakis D, Vicen-Bueno R. Sensors to Increase the Security of Underwater Communication Cables: A Review of Underwater Monitoring Sensors. *Sensors*. 2020;20(3):737.
- Been R, Hughes DT, Vermeij A. Heterogeneous underwater networks for ASW: technology and techniques. 2008 Jun [cited 2023 Jun 14]. Report No.: NURC-PR-2008-001. Available from: <https://openlibrary.cmre.nato.int/handle/20.50.0.12489/632>
- Jamieson A. *The Hadal Zone: Life in the Deepest Oceans*. Illustrated edition. Cambridge, United Kingdom ; New York: Cambridge University Press; 2015.
- Cong Y, Gu C, Zhang T, Gao Y. Underwater robot sensing technology: A survey. *Fundam Res*. 2021;1(3):337-45.
- Awan KM, Shah PA, Iqbal K, Gillani S, Ahmad W, Nam Y. Underwater Wireless Sensor Networks: A Review of Recent Issues and Challenges. *Wirel Commun Mob Comput*. 2019; 2019:e6470359.
- Menzel A. Maritime Security and the Blue Economy. In: *Routledge Handbook of Maritime Security*. Routledge; 2022.
- Greenacre M, Blasius J. Multiple correspondence analysis and related methods. 2006.
- Reimers N, Gurevych I. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks [Internet]. arXiv; 2019 [cited 2023 Aug 3]. Available from: <http://arxiv.org/abs/1908.10084>
- Liu Y, Ott M, Goyal N, Du J, Joshi M, Chen D, *et al.* RoBERTa: A Robustly Optimized BERT Pretraining Approach [Internet]. arXiv; 2019 [cited 2023 Aug 3]. Available from: <http://arxiv.org/abs/1907.11692>
- Roberts ME, Stewart BM, Tingley D. *stm*: An R package for structural topic models. *J Stat Softw*. 2019;91:1-40.
- Blei DM, Ng AY, Jordan MI. Latent Dirichlet allocation. *J Mach Learn Res*. 2003;3(4):993-1022.
- Ambrosino A, Cedrini M, Davis JB, Fiori S, Guerzoni M, Nuccio M. What topic modeling could reveal about the evolution of economics. *J Econ Methodol*. 2018;25(4):329-48.
- Figuerola CG, García Marco FJ, Pinto M. Mapping the evolution of library and information science (1978-2014) using topic modeling on LISA. *Scientometrics*. 2017;112(3):1507-35.
- Ryan R, Davis-Kean P, Bode L, Krüger J, Mneimneh Z, Singh L. Parenting online: analyzing information provided by parenting-focused Twitter accounts. *Atl J Commun*. 2022;0(0):1-17.
- Mosallaei S, Rad M, Schiffauerova A, Ebadi A. Discovering the evolution of artificial intelligence in cancer research using dynamic topic modeling. *COLLNET J Scientometr Inf Manag*. 2021;15(2):225-40.
- Gurcan F, Cagiltay NE. Exploratory Analysis of Topic Interests and Their Evolution in Bioinformatics Research Using Semantic Text Mining and Probabilistic Topic Modeling. *IEEE Access*. 2022;10:31480-93.
- Ebadi A, Xi P, Tremblay S, Spencer B, Pall R, Wong A. Understanding the temporal evolution of COVID-19 research through machine learning and natural language processing. *Scientometrics*. 2021;126(1):725-39.
- Lucas C, Nielsen RA, Roberts ME, Stewart BM, Storer A, Tingley D. Computer-assisted text analysis for comparative politics. *Polit Anal*. 2015;23(2):254-77.
- Maskeri G, Sarkar S, Heafield K. Mining business topics in source code using latent dirichlet allocation. In: *In Proceedings of the 1st India software engineering conference*. 2008. p. 113-20.

27. Ebadi A, Tremblay S, Goutte C, Schiffauerova A. Application of machine learning techniques to assess the trends and alignment of the funded research output. *J Informetr.* 2020;14(2):101018.
28. Churchill R, Singh L. The Evolution of Topic Modeling. *ACM Comput Surv.* 2022; 54(10s):215: 1-215:35.
29. Röder M, Both A, Hinneburg A. Exploring the Space of Topic Coherence Measures. In: Proceedings of the Eighth ACM International Conference on Web Search and Data Mining - WSDM '15. ACM Press; 2015. p. 399-408. Available from: <http://dl.acm.org/citation.cfm?doid=2684822.2685324>
30. Raffel C, Shazeer N, Roberts A, Lee K, Narang S, Matena M, *et al.* Exploring the limits of transfer learning with a unified text-to-text transformer. *J Mach Learn Res.* 2020; 21(1): 140: 5485-140:5551.
31. Bornmann L, Mutz R. Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references. *J Assoc Inf Sci Technol.* 2015;66(11):2215-22.
32. Callon M, Rip A, Law J, editors. Mapping the Dynamics of Science and Technology: Sociology of Science in the Real World. 1986th edition. Palgrave Macmillan; 1986.
33. Ding Y, Chowdhury GG, Foo S. Bibliometric cartography of information retrieval research by using co-word analysis. *Inf Process Manag.* 2001;37(6):817-42.
34. Bello O, Zeadally S. Internet of underwater things communication: Architecture, technologies, research challenges and future opportunities. *Ad Hoc Netw.* 2022;135:102933.
35. Mohsan SAH, Mazinani A, Othman NQH, Amjad H. Towards the internet of underwater things: a comprehensive survey. *Earth Sci Inform.* 2022;15(2):735-64.

Cite this article: Ebadi A, Auger A, Gauthier Y. Scientometric Insights into the Trends and Evolution of the Underwater Sensing Technologies and their Applications. *J Scientometric Res.* 2024;13(2):562-74.

APPENDIX A. Search query

Below is the search query utilized to retrieve publications related to underwater sensing technologies from Elsevier's Scopus (data were collected on September 14, 2022):

```
(((TITLE-ABS-KEY("underwater surveillance" OR "underwater threat detect*" OR "underwater anomal*" OR "underwater intrusion*" OR ("underwater sensor*" W/5 (surveillance OR threat* OR network*))) OR TITLE-ABS-KEY((underwater OR "under water" OR undersea OR "under sea" OR subsea OR sub-sea OR "under ocean" OR sub-ocean) AND (object OR objects) AND ("human made" OR manmade OR man-made OR "man made") AND (detect* OR sensor OR sensing))) AND PUBYEAR > 1989) OR ((TITLE-ABS-KEY((underwater OR "under water" OR undersea OR "under sea" OR subsea OR sub-sea OR "under ocean" OR sub-ocean) W/2 (surveillance OR target OR threat OR sensor OR intru* OR sensors OR sensing OR detect* OR militar*)) AND TITLE-ABS-KEY(submarine* OR anti-submarine* OR antisubmarine OR threat* OR collision* OR countermeasure* OR bomb* OR mine OR mines OR anomal* OR depth-charge* OR terroris* OR "magnetic object*" OR ferromagnetic OR "explosive device*" OR ordnance* OR weapon*) AND NOT TITLE-ABS-KEY(cable OR pipelines OR bird OR medicine OR threatened)) AND PUBYEAR > 1989) OR ((TITLE-ABS-KEY(submarine W/2 (detect* OR sensing OR sensor OR ident* OR track* OR hunt*) OR antisubmarine OR anti-submarine OR "anti submarine") AND NOT TITLE-ABS-KEY(cable* OR pipe* OR landslide* OR groundwater OR seismic OR earthquake*)) AND PUBYEAR > 1989) OR ((TITLE-ABS-KEY("underwater detection" OR "underwater surveillance" OR (underwater W/2 intru*) OR (underwater PRE/1 tracking)) AND PUBYEAR > 1989) OR ((AUTHKEY(underwater W/2 (image OR imaging))) OR (TITLE(underwater W/2 (image OR imaging)) AND PUBYEAR > 1989))) OR (TITLE-ABS-KEY(("under ice" OR (under PRE/1 ice) OR (under PRE/1 snow)) W/5 (detect* OR identif* OR recogni* OR track* OR sensing OR sensor OR imaging)) AND PUBYEAR > 1989) OR (TITLE-ABS-KEY("underwater object*" W/2 (detect* OR
```

```
identif* OR recogni* OR track* OR sensing)) AND PUBYEAR > 1989) OR ((TITLE("underwater acoustic sensor network*" OR uasn* OR "underwater wireless sensor network*" OR uwsn*) OR AUTHKEY("underwater acoustic sensor network*" OR uasn* OR "underwater wireless sensor network*" OR uwsn*)) AND PUBYEAR > 1989)) OR ((((((TITLE-ABS-KEY("underwater surveillance" OR "underwater threat detect*" OR "submarine sensing" OR "submarine identification" OR "submarine detection" OR "submarine tracking" OR "submarine hunting") OR TITLE-ABS-KEY("underwater anomal*" OR "underwater intrusion*" OR ("underwater sensor*" W/2 (surveillance OR threat* OR network*))) OR TITLE-ABS-KEY((underwater OR "under water" OR undersea OR "under sea" OR subsea OR sub-sea OR "under ocean" OR sub-ocean OR maritime) AND object* AND (manmade OR man-made OR "man made") AND (detect* OR sens*)) AND SUBJAREA(mult OR ceng OR chem OR comp OR eart OR ener OR engi OR envi OR mate OR math OR phys OR mult OR arts OR busi OR deci OR econ OR psyc OR soci) AND PUBYEAR > 1989) OR ((TITLE-ABS-KEY((underwater OR "under water" OR undersea OR "under sea" OR subsea OR sub-sea OR "under ocean" OR sub-ocean OR maritime) W/5 (surveillance OR sensor OR sensors OR sensing OR detection OR militar*)) AND TITLE-ABS-KEY((submarine* OR anti-submarine* OR antisubmarine OR threat* OR collision* OR countermeasure* OR bomb* OR mine OR mines OR anomal* OR depth-charge* OR terroris* OR "magnetic object*" OR ferromagnetic OR "explosive device*" OR ordnance* OR weapon*)) AND SUBJAREA(mult OR ceng OR chem OR comp OR eart OR ener OR engi OR envi OR mate OR math OR phys OR mult OR arts OR busi OR deci OR econ OR psyc OR soci) AND PUBYEAR > 1989)) OR (TITLE-ABS-KEY((underwater OR "under water" OR undersea OR "under sea" OR subsea OR sub-sea OR "under ocean" OR sub-ocean OR maritime) W/15 (surveillance OR detect* OR monitor* OR militar* OR terroris* OR warfare* OR submarine* OR anti-submarine* OR antisubmarine OR threat* OR collision* OR countermeasure* OR bomb* OR mine OR mines OR anomal* OR depth-charge* OR "magnetic object*" OR ferromagnetic OR explosive* OR
```

ordnance* OR weapon*) W/15 (sensor OR sensors OR sensing)) AND SUBJAREA(mult OR agri OR bioc OR immu OR neur OR phar OR mult OR ceng OR chem OR comp OR eart OR ener OR engi OR envi OR mate OR math OR phys OR mult OR arts OR busi OR deci OR econ OR psyc OR soci) AND PUBYEAR > 1989)) AND NOT((TITLE-ABS-KEY((oil OR gas) W/5 (spill* OR leak*)) OR TITLE-ABS-KEY(aircraft) OR TITLE-ABS-KEY(earthquake* OR tsunami*)) AND SUBJAREA(mult OR medi OR nurs OR vete OR dent OR heal OR mult OR ceng OR chem OR comp OR eart OR ener OR engi OR envi OR mate OR math OR phys OR mult OR arts OR busi OR deci OR econ OR psyc OR soci) AND PUBYEAR > 1989)) AND NOT((TITLE(pipeline OR "pipe line*") OR TITLE-ABS-KEY(groundwater) OR TITLE-ABS-KEY(soil)) AND SUBJAREA(mult OR medi OR nurs OR vete OR dent OR heal OR mult OR ceng OR chem OR comp OR eart OR ener OR engi OR envi OR mate OR math OR phys OR mult OR arts OR busi OR deci OR econ OR psyc OR soci) AND PUBYEAR > 1989)) AND NOT((KEY(pipeline OR "pipe line*") AND SUBJAREA(mult OR medi OR nurs OR vete OR dent OR heal OR mult OR ceng OR chem OR comp OR eart OR ener OR engi OR envi OR mate OR math OR phys OR mult OR arts OR busi OR deci OR econ OR psyc OR soci) AND PUBYEAR > 1989))) OR ((TITLE("underwater imaging") OR AUTHKEY("underwater imaging"))) AND (LIMIT-TO(PUBYEAR,2022) OR LIMIT-TO(PUBYEAR,2021) OR LIMIT-TO(PUBYEAR,2020) OR LIMIT-TO(PUBYEAR,2019) OR LIMIT-TO(PUBYEAR,2018) OR LIMIT-TO(PUBYEAR,2017) OR LIMIT-TO(PUBYEAR,2016) OR LIMIT-TO(PUBYEAR,2015) OR LIMIT-TO(PUBYEAR,2014) OR LIMIT-TO(PUBYEAR,2013) OR LIMIT-TO(PUBYEAR,2012) OR LIMIT-TO(PUBYEAR,2011) OR LIMIT-TO(PUBYEAR,2010) OR LIMIT-TO(PUBYEAR,2009) OR LIMIT-TO(PUBYEAR,2008) OR LIMIT-TO(PUBYEAR,2007) OR LIMIT-TO(PUBYEAR,2006) OR LIMIT-TO(PUBYEAR,2005) OR LIMIT-TO(PUBYEAR,2004) OR LIMIT-TO(PUBYEAR,2003) OR LIMIT-TO(PUBYEAR,2002) OR LIMIT-TO(PUBYEAR,2001) OR LIMIT-TO(PUBYEAR,2000) OR LIMIT-TO(PUBYEAR,1999) OR LIMIT-TO(PUBYEAR,1998) OR LIMIT-TO(PUBYEAR,1997) OR LIMIT-TO(PUBYEAR,1996) OR LIMIT-TO(PUBYEAR,1995) OR LIMIT-TO(PUBYEAR,1994) OR LIMIT-TO(PUBYEAR,1993) OR LIMIT-TO(PUBYEAR,1992) OR LIMIT-TO(PUBYEAR,1991) OR LIMIT-TO(PUBYEAR,1990))