



OPEN Rendering the European neutron research landscape

Evgenii Velichko^{1✉}, Hartmut Abele², David J. Barlow³, Antonio Benedetto^{4,5}, Stefano Deledda⁶, Lambert van Eijck¹, Maria Teresa Fernandez-Diaz⁷, Marc Janoschek^{8,9}, Maths Karlsson¹⁰, Kim Lefmann¹¹, Natalie Malikova¹², Maria P. M. Marques^{13,14}, Konstantina Mergia¹⁵, Jörg Pieper¹⁶, Henrik M. Rønnow^{11,17}, Jan Šaroun¹⁸, Astrid Schneidewind¹⁹, Frank Schreiber²⁰, Kristiaan Temst^{21,22}, Max Wolff²³, Wojciech Zając²⁴ & Marco Zanatta²⁵

Neutrons, owing to their unique properties, serve as indispensable probes for investigating the structure and dynamics of materials across various length scales. The scientific community utilizing neutron research infrastructures encompasses a diverse range of disciplines, making it challenging to quantify its scientific and societal impact. To address this challenge, we apply Natural Language Processing (NLP) and machine learning techniques to analyze the scientific output of the European neutron science community. Leveraging open-source software toolkits, our method allows for the quantitative assessment of community evolution and research focus. Our analysis reveals consistent growth in the neutron community despite a reduction in sources, underscoring the enduring significance of neutron methods in scientific research. Furthermore, an increase in unique authors and an even distribution of publications across diverse scientific topics highlight the community's interdisciplinary nature and collaborative spirit. While this study emphasizes neutron scattering, our methodology holds promise for a broad range of scientific communities reliant on Large Research Infrastructures (LRIs), offering opportunities for collaboration, optimization of experimental approaches, and informed decision-making by governmental and funding bodies.

Neutrons - by virtue of their electrical neutrality and the nuclear character of their interaction with matter - provide a non-invasive and non-destructive means to probe the structure and dynamics of materials from atomic to mesoscopic or even macroscopic length scales. Thanks to their magnetic dipole, they can also probe magnetism. Neutrons can be employed under extreme and versatile environments, and generally are considered a unique and indispensable tool to study all forms of condensed matter. Instrumentation for neutron research is mostly available at large scale facilities, where each neutron source serves several beamlines with instrumentation dedicated to serve specific scientific fields. Inherently, such facilities serve a broad range of scientific applications ranging from imaging at centimeter to micrometer scale, down to macromolecular and sub-atomic distances,

¹Department of Radiation Science and Technology, Faculty of Applied Sciences, TU Delft, Delft, Netherlands. ²TU Wien Atominstytut, Wien, Austria. ³Division of Pharmacy & Optometry, University of Manchester, Manchester, UK. ⁴School of Physics, University College Dublin, Dublin, Ireland. ⁵Conway Institute of Biomolecular and Biomedical Research, University College Dublin, Dublin, Ireland. ⁶Department for Hydrogen Technology, Institute for Energy Technology, Kjeller, Norway. ⁷Institut Laue-Langevin, Grenoble, France. ⁸Laboratory for Neutron and Muon Instrumentation, Paul Scherrer Institute, Villigen, Switzerland. ⁹Physik-Institut, Universität Zürich, Zürich, Switzerland. ¹⁰Department of Chemistry and Chemical Engineering, Chalmers University of Technology, Göteborg, Sweden. ¹¹Nanoscience Center, Niels Bohr Institute, University of Copenhagen, Copenhagen, Denmark. ¹²Laboratory of Physical Chemistry of Electrolytes and Interfacial Nanosystems (PHENIX), Sorbonne Université, CNRS, Paris, France. ¹³Molecular Physical-Chemistry R&D Unit, Department of Chemistry, University of Coimbra, Coimbra, Portugal. ¹⁴Department of Life Sciences, University of Coimbra, Coimbra, Portugal. ¹⁵Institute of Nuclear and Radiological Science and Technology, Energy and Safety, National Centre for Scientific Research "Demokritos", Athens, Greece. ¹⁶Institute of Physics, University of Tartu, Tartu, Estonia. ¹⁷Laboratory for Quantum Magnetism (LQM), Institute of Physics, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland. ¹⁸Nuclear Physics Institute of the CAS, Řež, Czech Republic. ¹⁹JCNS at MLZ, FZ Jülich, Garching, Germany. ²⁰Institut für Angewandte Physik, Universität Tübingen, Tübingen, Germany. ²¹Quantum Solid State Physics division, KU Leuven, Leuven, Belgium. ²²Interuniversity Microelectronics Center imec, Leuven, Belgium. ²³Division for Materials Physics, Department of Physics and Astronomy, Uppsala University, Uppsala, Sweden. ²⁴Institute of Nuclear Physics, Polish Academy of Sciences, Krakow, Poland. ²⁵Department of Physics, University of Trento, Trento, Italy. ✉email: evgentudelft@gmail.com

and femtosecond to microsecond dynamics. The use of neutrons as scientific probes has evolved from solid state physics and chemical crystallography to a much broader range of topics, including soft matter, nanotechnology, biology, cultural heritage and engineering^{1,2}. As a consequence of the breadth of topics covered, the scientific and societal impact of the community of ‘neutron scientists’ is significant but difficult to quantify and visualize.

To address this challenge, we analyze the scientific output of the European neutron science community using Natural Language Processing (NLP) and machine learning, to render an overview of the rather broad European community of scientists that use the neutron research infrastructures for their scientific research. This study is primarily focused on the specific results for neutron scattering, while also demonstrating the novelty and applicability of our machine learning approach. Through this novel approach we are able to showcase the evolution and distribution of the community and the main foci of research in a quantitative manner. The outcome of this approach is shown to be consistent with the findings reported by others^{3–5}.

The approach to use open-source software to analyze a scientific community through its publications originated from our participation as European Neutron Scattering Association (ENSA¹), in the Horizon2020 Brightness2 project⁶. The development of the analysis tools served to describe the needs of the community for the long-term sustainability of the European Spallation Source (ESS⁷), . We extended these tools to use semi-supervised machine learning to quantitatively describe the community and their scientific focus. While we are aware that there are other approaches to this type of analysis, we argue that our method allows us to render an overview of a community and its impact in an unbiased and quantitative manner, even when the community is rather disperse, broad and heterogeneous.

Method

Data collection

The metadata of the publications were obtained from the Scopus database using the Scopus Search API, facilitated by the `pybliometrics`⁸ python package. To compile a comprehensive database of publications related to neutron research, a straightforward term search was conducted on titles, keywords, and abstracts using the term “neutron” (“TITLE-ABS-KEY(neutron)”). To validate the relevance of the retrieved entries, two additional databases of publication metadata were collected, specifically including entries associated with the ILL affiliation ID (60007109) and the ISIS Neutron and Muon Source affiliation ID (60001724). On the basis of the collected metadata, we created a Venn diagram to validate our search approach.

Data filtering – supervised machine learning

To filter out publications containing the term “neutron” that were deemed irrelevant to the neutron scattering community (e.g., references to ‘neutron stars’), we employed a supervised machine learning approach. This process aimed to classify publications into two categories: “relevant” and “irrelevant” to the neutron scattering community.

For the development of a classification model, we enlisted neutron scientists from the European Neutron Scattering Association (ENSA) to manually label a dataset of 13,139 publications from the years 2021 and 2022 as either “from the community (relevant)” or “not from the community (irrelevant).” These community representatives were provided with the metadata of each publication in the dataset. The representatives also could inspect the full text of the publication via its DOI, though the machine learning algorithm itself only utilized metadata.

The labeled dataset was then split into two subsets: a training set, which contained 90% of the labeled publications (11,779 entries), and a test set, which contained the remaining 10% (1,360 entries). We used the training subset to train a classification model, employing a stochastic gradient descent classifier⁹ from the `scikit-learn` Python package¹⁰. The model was evaluated on the test subset and achieved an accuracy of 89%.

After training, we applied this model to classify the entire corpus of “neutron” publications, which had been initially generated by a Scopus query based on the single term “neutron.” Using this supervised machine learning model allowed us to reduce the corpus size by removing entries classified as “irrelevant” to the neutron scattering community. Only publications deemed “relevant” by the model were included in the subsequent analysis.

Statistics

From the filtered database, the main publication trends were extracted, including the number of publications per year, the average number of (co-)authors and (co-)affiliations per publication per year, as well as the total number of (co-)authors and (co-)affiliations per year. For our analysis, we made no distinction between the main author, corresponding author or co-authors. Every person included within the metadata, was thus considered ‘author’ for the purposes of the subsequent analyses.

NLP topic modelling – unsupervised machine learning

An unsupervised machine learning technique called Latent Dirichlet Allocation (LDA)¹¹ was employed to identify common topics frequently occurring in the neutron publications. The number of topics used in the LDA analysis ranged from 2 to 50, and the optimal number of topics was determined based on the extent of topic separation and avoidance of excessive fragmentation.

For the analysis, the titles, keywords, and abstracts of the publications were combined into a single text entry. These combined text entries were then tokenized and filtered using the `NLTK`¹² python package. To identify a non-topic-specific vocabulary that appeared frequently across the entire corpus, an initial iteration of LDA was performed using a standard list of English stopwords and punctuation characters provided by the `NLTK` package. The identified vocabulary was then added to the list of stopwords for subsequent modeling, and further details can be found in Appendix 1.

All the tokenized and filtered text entries were transformed into vectors in an n-dimensional token space using the count vectorizer provided by the scikit-learn python package. LDA was subsequently applied to these vectorized texts, and the resulting topics were visualized using the LDAvis¹³ package.

Based on the identified topics, each text entry was assigned a percentage score indicating its alignment with each of the identified topics.

Description of trends in the European neutron science community

The data of individual authors, including their affiliations, city, and country, were extracted from the publications database. To understand global trends in the number of authors involved in neutron research, the unique Scopus author IDs were classified into six world regions aiming to provide insights into the dynamics of the neutron community.

Furthermore, the authors were categorized into two groups: “new,” representing authors appearing in the database for the first time at a specified year, and “old,” representing authors who were already present in the database prior to that year.

To visualize the neutron community, the authors’ locations (country and city, when available) were geocoded using the OpenStreetMap¹⁴ Nominatim API. However, due to inconsistencies in the author data, such as missing or mismatched city and country information, an adapted strategy was employed. Authors without a specified country in the publications metadata were excluded from the visualization. For authors with a provided country, their longitude and latitude coordinates were initially determined based on the country. In cases where automatic geocoding failed, a manual inspection was conducted, and the names of corresponding territories were updated (e.g., replacing “Libyan Arab Jamahiriya” with “Libya” and “German Democratic Republic” with “Germany”). Subsequently, for authors with both city and country information available, their coordinates were adjusted to the specific city. All authors were then represented as author-density distribution as a “heatmap” on a world map using the folium¹⁵ Python library.

Results & discussion

Data collection

The publications metadata were collected from the Scopus database via the Scopus Search API with the help of pybliometrics python package⁸. An API search for the keyword “neutron” yielded 320,357 hits, covering the years 1922–2022. In order to have a reliable benchmark, we also searched the publications metadata for two large European neutron sources, namely, the Institut Laue Langevin (ILL, Grenoble, France) and ISIS Neutron and Muon Source (Rutherford Appleton Laboratories, Didcot, UK). The search yielded 15,898 entries in the period 1968–2022 and 9149 in the period 1987–2022 for ILL and ISIS Neutron and Muon Source respectively (We note that a considerable ambiguity exists for the links between publications and affiliations. For example, publications linked to ISIS Neutron and Muon Source could have Rutherford Appleton Laboratory (RAL), or Science and Technology Facilities Council (STFC) in their list of affiliations instead of ISIS Neutron and Muon Source. Therefore, we are convinced that not all the publications originating from ISIS Neutron and Muon Source are present in our results, but we argue that our search delivered a representative fraction of these. On the other hand, the ILL neutron source has no alternative names or IDs, and therefore we suppose that our search results should include nearly all the publications containing ILL in their affiliations lists). Figure 1 represents a Venn diagram of the publications found on Scopus by aforementioned search queries. It can be seen that the majority of the publications originating from ILL and ISIS neutron sources are present in the search result for

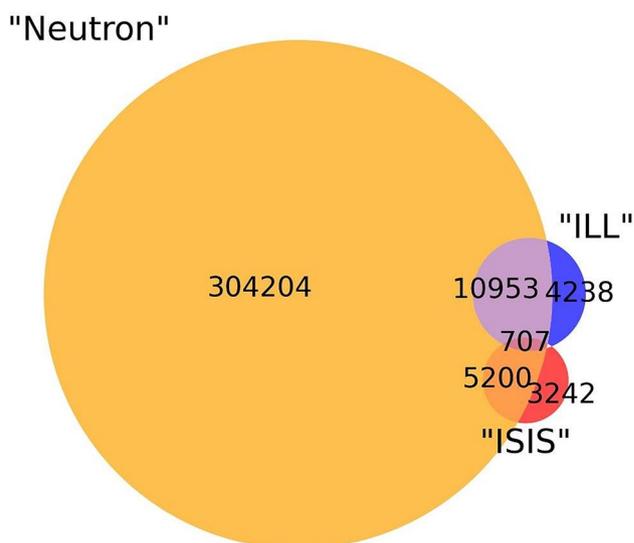


Fig. 1. A Venn diagram of the publications entries found in Scopus. It visualizes that a search on the term “neutron” leads to capturing of about 65% of the entries which have ISIS Neutron and Muon Source or ILL sources in their lists of affiliation ID’s.

$N=1360$	Predicted: ENSA	Predicted: non-ENSA
Actual: ENSA	True Positive (TP=739)	False Negative (FN=55)
Actual: non-ENSA	False Positive (FP=90)	True Negative (TN=476)
Overall Accuracy: $(TP + TN)/(TP + TN + FP + FN) = 0.89$		

Table 1. Supervised machine learning accuracy results on the test data subset of 1360 publications. The algorithm identified 739 + 90 publications to be ENSA-related, of which 739 were actually identified by the scientists to be ENSA-related. As such the combinations of True/False and Positive/Negative yields an algorithm accuracy of 89%.

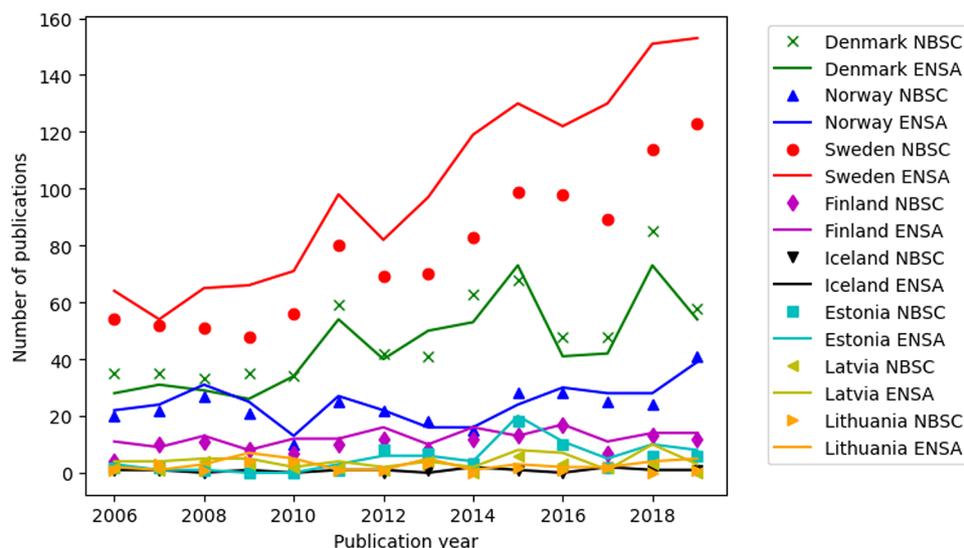


Fig. 2. A comparison of numbers of neutron publications for Nordic and Baltic Neutron Scattering Communities (NBSC) based on a Nordic report using manual classification⁴ (markers, NBSC) and our database (lines, ENSA).

the “neutron” keyword. Moreover, all of the publications originating from both affiliations simultaneously are captured in the search result for the “neutron” keyword.

Data filtering – supervised machine learning

The initial Scopus query based solely on the term “neutron” yielded a corpus much larger than the actual body of publications produced by the neutron scattering research community. Filtering out publications with terms such as “neutron star” naturally reduced the corpus size, but a more comprehensive filtering approach was necessary to avoid bias from overly restrictive criteria.

By employing the machine learning model trained on community-labeled data, we were able to classify publications based on the entire vocabulary of each entry’s metadata, rather than relying on a few specific terms. The accuracy of this filtering approach was evaluated based on the overlap between the model’s classifications and those provided by the neutron scientists. As shown in Table 1, the trained model achieved an accuracy of 89%, which we deemed sufficient for identifying publications relevant to the neutron scattering community.

Another indication of the ‘accuracy’ of our approach is depicted in Fig. 2, where the publication output of the Nordic neutron scattering community is counted by a manual analysis method (reading all abstracts, Lefmann⁴) and overlaid with similar numbers generated from our analysis. The numbers for most countries are reproduced well, while publications from Sweden show a 20% discrepancy.

Applying this classification model to the entire corpus of publication metadata collected from Scopus (320,357 entries), selecting only the entries classified as “relevant” to the community, and restricting the publication years to 1930–2020, reduced the number of entries in our publication dataset from 320,357 to 121,731. Detailed descriptions of the vocabularies relevant to the topics modeled for the filtered-out publications and those kept for further analysis are provided in Appendices B and C, respectively.

Statistics

In order to focus our analysis on neutron research involving the European neutron community, we split the filtered dataset into two subsets: one including at least one author with an affiliation in a European country (70,830 entries), and the other without any authors with European affiliations (50,901 entries).

The main data of our analysis are displayed in Fig. 3 and show that the number of publications of the European community rapidly increased during the last century, while being rather steady over the last 20 years. In contrast,

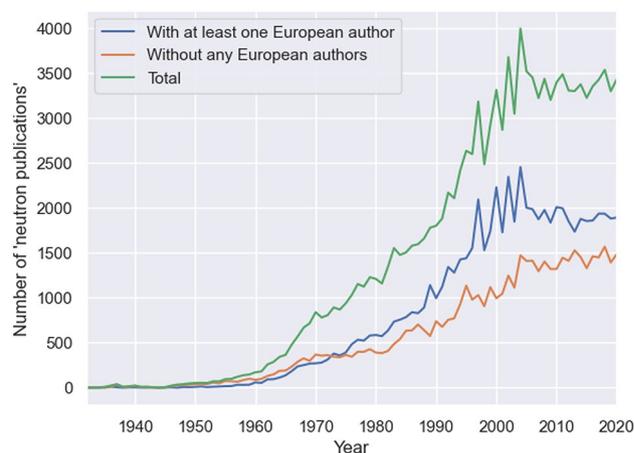


Fig. 3. The evolution of the number of ‘neutron publications’ per year, as derived from the filtered corpus of collected publications. More than half of the publications have at least one European author.

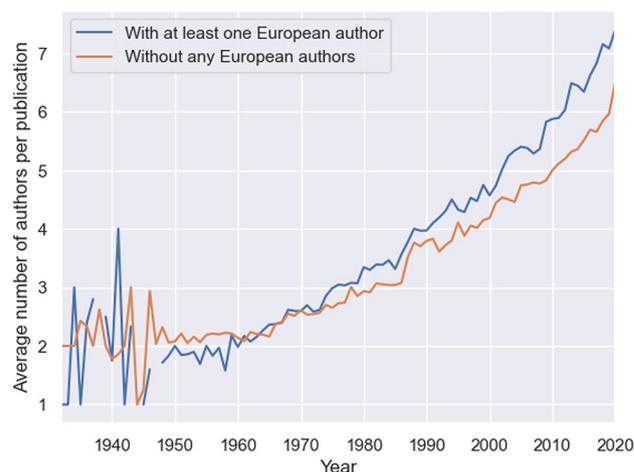


Fig. 4. Number of authors per publication per year, both for non-European and for European neutron publications.

the non-European publications show a steadily increasing trend, with the European researchers being involved in slightly over half of all publications world-wide. Strong variations (per year) occur in the decade around 2000, specifically in Europe. The reason for this is that many of the articles published over this period were associated with the bi-annual proceedings of the European Conference on Neutron Scattering (ECNS) and International Conference on Neutron Scattering (ICNS).

Over the same time span, the mean number of authors per publication has increased by a factor of 3.5 (Fig. 4). This is an indirect demonstration of how the community has evolved through the formation of larger, multidisciplinary teams that combine the use of neutrons with other scientific methods.

NLP topic modelling – unsupervised machine learning

The analysis of the metadata vocabulary enables the classification of published works into topics using NLP. This topic modeling is unbiased regarding the text’s meaning, though the authors select the most concise topic names or labels. For instance, the most relevant terms for topic 10 include ‘propagation vector,’ ‘magnetic phase diagram,’ ‘external magnetic,’ ‘metamagnetic,’ ‘zero field,’ ‘noncollinear,’ and ‘spin ice.’ Consequently, the topic was labeled ‘Magnetism’ for conciseness. A complete description of the modeled topics with the most relevant and most frequent terms can be found in the Appendix C.

Figure 5 shows how the relative output varies between the topics over the entire time span, which was dominated by magnetism, excitations, macroscopic structures and chemical composition in the late fifties, while the most recent 5 years show how neutron methods are applied rather evenly distributed over all 10 topics.

Description of trends in European neutron science community

Through the assigned affiliation of the authors in the corpus we were able to render the community distribution as geographical “heatmaps”, as shown in Fig. 6. Although there are (and were) only a few European neutron

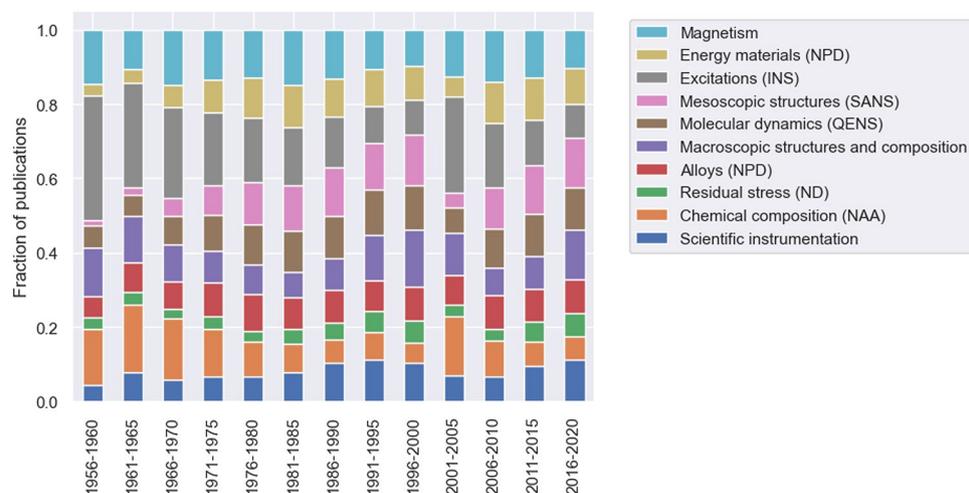


Fig. 5. Relative topic distribution of the ‘neutron publications’ with at least one European author. The names of the 10 topics coarsely describe the scientific field. When the NLP modelling identified a topic predominantly on the vocabulary that directly related to a specific applied neutron method, the topic name includes the method in brackets: Neutron Powder Diffraction (NPD), Inelastic Neutron Scattering (INS), Small Angle Neutron Scattering (SANS), Quasi-elastic Neutron Scattering (QENS), Neutron Diffraction (ND), and Neutron Activation Analysis (NAA). Other methods are present, but less strongly related to a single topic.

sources^{2,16} operational, over the last decades, the scientific community is now distributed much more broadly over European academia and industries. Because most neutron sources operate as international user facilities, even countries without a neutron source (e.g. countries like Italy and Spain) form a considerable fraction of the community.

The combination of NLP topic modelling and affiliations associated with any publication, allows to map the ‘topic density’ as geographical heatmaps, as shown in Fig. 7. For this visualization only publications, which are at least 50% aligned with one of the modelled topics were selected and attributed into the respective set. There is thus no double count across topics (with each publication appearing only in one of the topics). However, each affiliation involved in the publication is assigned a unit value for the publication. Therefore, each publication could be counted multiple times within the same topic. The total number of publications originating from each geographical location is expressed as colour intensity in Fig. 7. The heatmaps show a rather homogeneous distribution of scientific topics over the continent, indicating that each nation has scientists carrying out research on all topics. This map could be of great help for scientists looking for collaborations within or outside of their field of research.

Figure 8 shows the growth of the neutron community and the leading role of Europe, while Figs. 9 and 10 illustrate how a considerable fraction of this community is new to the field. Figure 10 demonstrates that about one third of the unique authors from Europe publish for the first time each year. This fraction indicates a large potential for the growth if the European Neutron Research community. The growth of this fraction is related to the fact that more recent publications tend to have more authors (Fig. 4) and that visiting scientists at the neutron user facilities nowadays often include researchers with little or no previous experience of neutron scattering research. This, at the same time, showcases that expertise in neutron scattering methods concentrated at European large-scale neutron facilities is crucial for the continuous growth of the community.

Conclusions

Leveraging published research and open-source machine learning toolkits, we have developed a method for conducting bibliometric analyses based on Natural Language Processing (NLP) and supervised machine learning. While the primary focus of this study is on the neutron research community, the methodology presented is innovative and broadly applicable to various scientific fields. This innovative approach, geared towards quantitative trend identification within scientific communities, was successfully tested on the international neutron scattering community, with a specific focus on the European landscape.

In the European neutron scattering community, our method revealed noteworthy trends. Despite a reduction in neutron sources, the neutron community in Europe kept growing, and has maintained its publication output rate over the past two decades, emphasizing the enduring significance of the use of neutron methods in scientific research. The continuous rise in unique authors, particularly among newcomers, indicates sustained interest and positivity within the neutron research community. Furthermore, an even distribution of publications and authors across diverse scientific topics highlights the community’s interdisciplinary nature and collaborative ethos, positioning it favorably amid changes in the scientific landscape, including the integration of new sources like the European Spallation Source (ESS⁷), and compact sources^{17–19}.

Although we were able to gain remarkable insights into the structure and dynamics of the European neutron scattering community using this novel approach, there are several limitations to our method as applied in

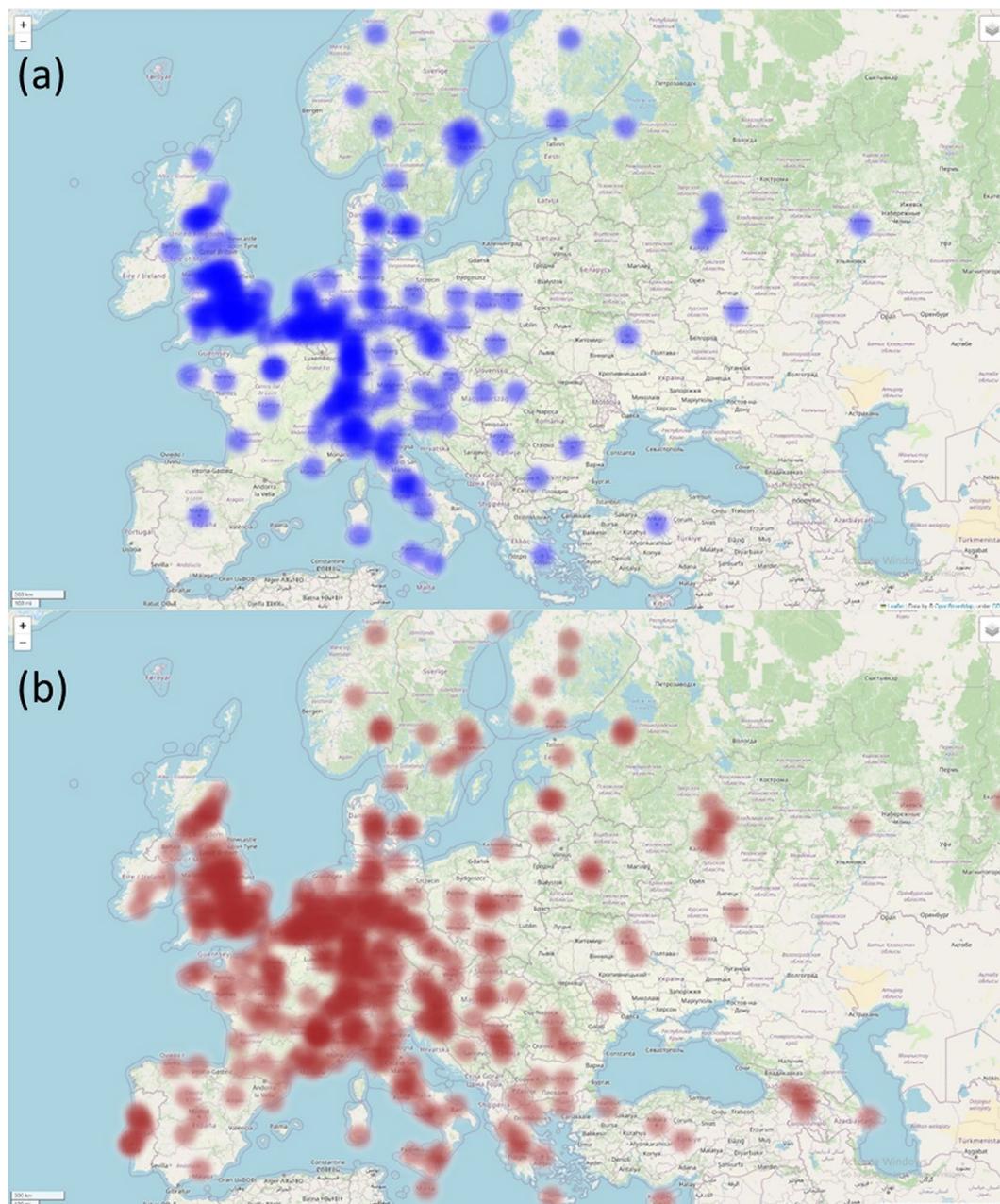
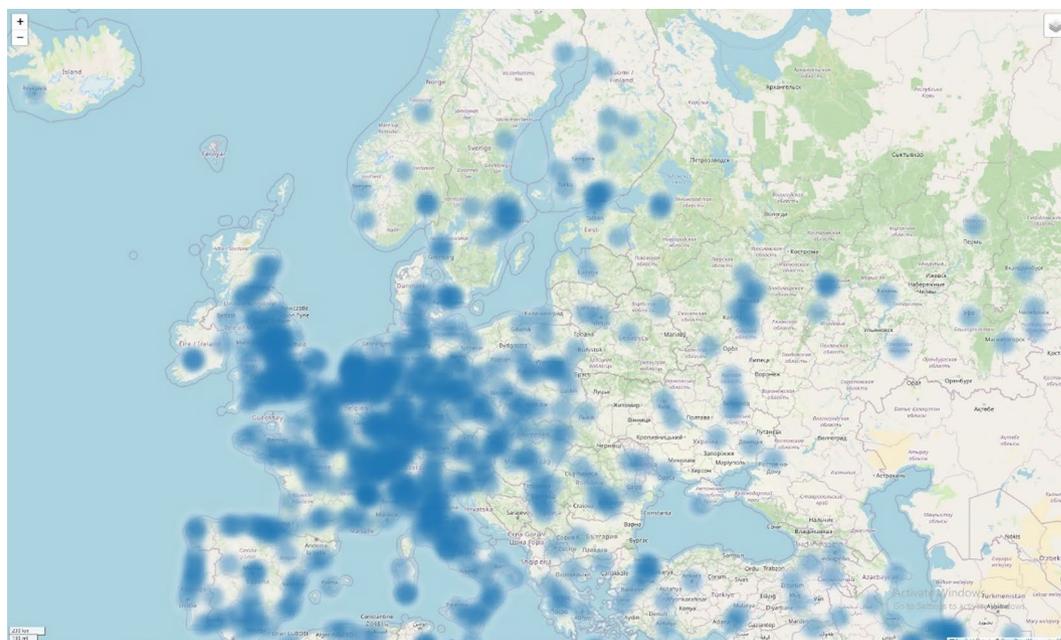


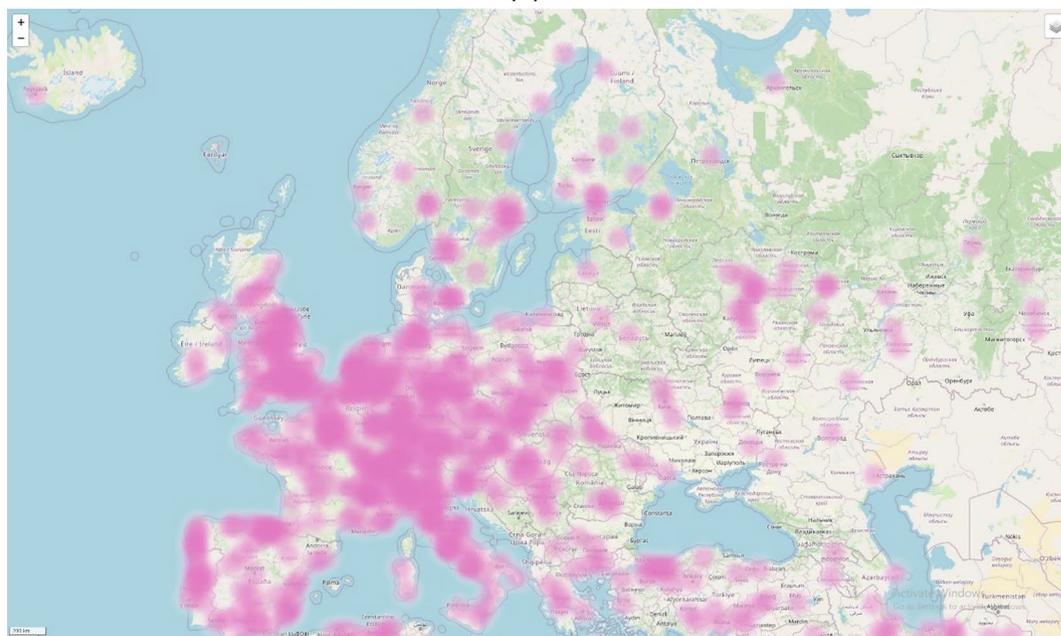
Fig. 6. The geographic distribution of the neutron science community depicted as heatmaps. (a) The heatmap for 1961–1970 is plotted (blue) above the (b) 2011–2020 community heatmap (red) to show how the community spread has evolved. [interactive version as supplementary information: [Europe_timesplit_map.html](#)]. The heatmaps are not normalized to population density. Map generated using the Python Folium package.

this work. Firstly, we used only publication metadata in our analysis and not the full texts of the publications. Extending the analysis to include the complete texts of publications would allow for a deeper examination of correlations between specific neutron techniques, the materials typically studied with these techniques, and other non-neutron techniques involved in such investigations. Furthermore, in our approach, we included all authors of each publication as being involved in neutron research. However, there is a clear distinction between users active at the neutron facility and other contributors to the research. A more detailed analysis of the roles of individual authors could be achieved by incorporating the full texts of publications in the analysis. Moreover, there is significant interest in similar analyses from the synchrotron user community^{20,21}. We are confident that our method could be beneficial to them as well, and through collaboration, we can further develop and refine this approach.

The insights gained by the application of our method extend beyond neutron scattering, presenting a versatile methodology applicable to various scientific communities, especially those reliant on Large Research



(a)



(b)

Fig. 7. The geographical spread of publication density for the 2 of the 10 identified scientific topics: (a) Scientific instrumentation, (b) Mesoscopic structures (SANS). No correction is done for population density variations over the continent. [interactive version as supplementary information: [Europe_neutron_all_topics_map.html](#)].

Infrastructures (LRIs) such as Synchrotrons, Accelerators, and Telescopes. Individual researchers and groups can benefit by establishing collaborations across adjacent fields, optimizing experimental approaches. Research infrastructures can leverage the method to swiftly understand evolving community needs, aligning facilities with dynamic research requirements. Governmental and funding bodies can make informed decisions, identifying research directions aligned with global policies and initiatives. Additionally, the method facilitates the allocation of short-term funds to actively contributing research groups, underlining its potential as a valuable asset for diverse stakeholders shaping the scientific research landscape.

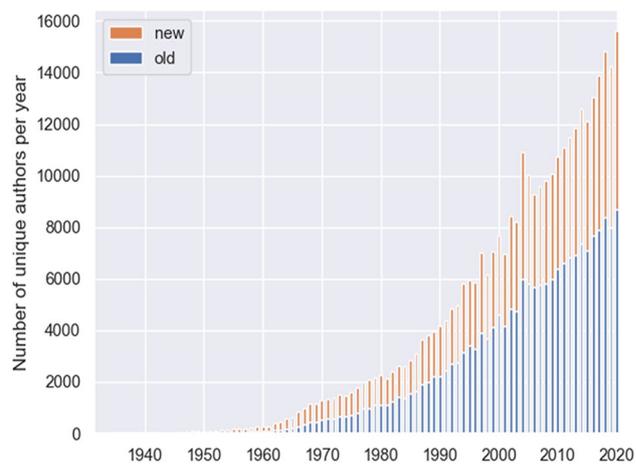


Fig. 8. Trend and distribution of the authors involved in neutron publications around the world. The total number of unique authors around the world is 150,421.

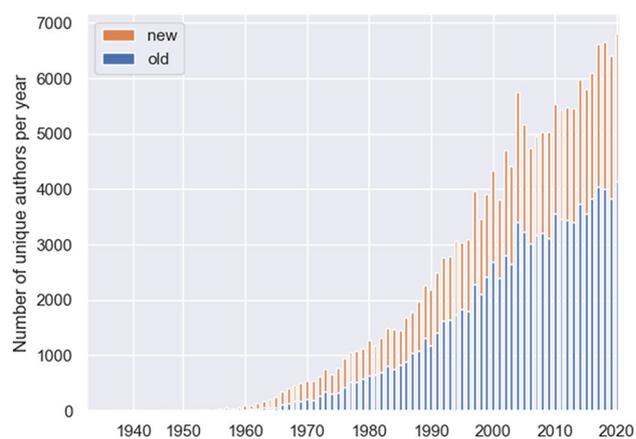


Fig. 9. The community of neutron scientists, separated into scientists that appear for the first time in a neutron publication ('new') and those that appear for a second time, or more ('old').



Fig. 10. The evolution of the number of European authors, indicating the number of authors that appear for the first time in the corpus, that year ('new'). The total number of unique authors from Europe is 71,311.

Data availability

The datasets used and/or analysed during the current study available from the corresponding author on reasonable request.

Received: 4 February 2024; Accepted: 24 January 2025

Published online: 17 February 2025

References

1. European Neutron Scattering Association. <https://neutrons-ensa.eu>.
2. League of Advanced Neutron Sources (LENS). <https://lens-initiative.org/>.
3. Gutberlet, T., Tunger, D., Zeitler, P. & Brückel, T. Do neutrons publish? A neutron publication survey, 2005–2015. *Neutron News*. **29** (2), 18–24. <https://doi.org/10.1080/10448632.2018.1514199> (2018).
4. Lindholm, M.K., Christensen, D.L. & Lefmann, K. *Nordic and Baltic Neutron Scattering Communities, 2006–2022 - A Bibliometric Study*. https://nnspl.nbi.ku.dk/the-community/Nordic_Baltic_Neutron_users_2022.pdf.
5. Hanna Barriga, M. et al. A bibliometric study on Swedish neutron users for the period 2006–2020. *Neutron News*. **32** (4), 28–33. <https://doi.org/10.1080/10448632.2021.1999147> (2021).
6. Horizon2020 grant agreement number 823867. <https://brightness.esss.se/>.
7. ESS. <https://europeanspallationsource.se/>.
8. Rose, M. E. & Kitchin, J. R. pybliometrics: Scriptable bibliometrics using a Python interface to Scopus. *SoftwareX* **10** (2019). <https://doi.org/10.1016/j.softx.2019.100263>.
9. Léon, B. Stochastic gradient learning in neural networks. In *Proceedings of Neuro-Nimes 91, EC2, Nimes, France* (1991). <http://leon.bottou.org/papers/bottou-91c>.
10. Pedregosa, F. et al. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **12** (2011). <http://scikit-learn.sourceforge.net>.
11. Blei, D. M., Ng, A. Y. & Jordan, M. I. January Lafferty, J. (ed.). Latent Dirichlet allocation. *J. Mach. Learn. Res.* **3**(4–5), 993–1022 (2003). <https://doi.org/10.1162/jmlr.2003.3.4-5.993>
12. Bird, S., Loper, E. & Klein, E. *Natural Language Processing with Python* (O'Reilly Media Inc., 2009).
13. Sievert, C. & Shirley, K. LDAvis: A method for visualizing and interpreting topics. In *Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces*. 63–70. (2014).
14. Stowell et al. *OpenStreetMap*. <https://www.openstreetmap.org/> (2020).
15. Python-visualization. *Folium*. <https://python-visualization.github.io/folium/> (2020).
16. ESFRI. <https://str-esfri.eu/working-groups/neutron-landscape-group>.
17. Carpenter, J. M. The development of compact neutron sources. *Nat. Rev. Phys.* **1**, 177–179. <https://doi.org/10.1038/s42254-019-0024-8> (2019).
18. Zimmer, M. et al. Demonstration of non-destructive and isotope-sensitive material analysis using a short-pulsed laser-driven epithermal neutron source. *Nat. Commun.* **13**, 1173. <https://doi.org/10.1038/s41467-022-28756-0> (2022).
19. Tomohiro Kobayashi, S., Ikeda, Y., Otake, Y., Ikeda, N. & Hayashizaki Completion of a new accelerator-driven compact neutron source prototype RANS-II for on-site use. In *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*. Vol. **994**. 165091. ISSN 0168–9002 (2021). <https://doi.org/10.1016/j.nima.2021.165091>.
20. NEPHEWS. <https://beamtime.eu/>.
21. Stankovski, M. & Farhad, A. P. Khotbehsara Adventures in Bibliography Part I: Uncovering the Global Status Quo (2024). <https://www.linxs.se/news/article-series-i-size-of-the-global-light-and-neutron-source-communities>.

Acknowledgements

This work was supported by the European Union's Horizon2020 project Brightness2 [<https://brightness.esss.se/>] (grant agreement 823867). We thank Prof. C. Alba-Simionesco for enabling ENSA participation to Brightness² that led to this research. We thank Prof. M. R. Johnson for valuable discussions on benchmarking our method to the databases of the Institut Laue-Langevin, Grenoble. We thank Dr. T. Gutberlet of FZ-Jülich for valuable discussions on our method in comparison to (his) earlier work.

Author contributions

The authors HA, DJB, AB, SD, LvE, MTFD, MJ, MK, KL, NM, MPMM, KM, JP, HR, JS, AS, FS, KT, MW, WZ, MZ are/were national delegates to the European Neutrons Scattering Association (ENSA) and contributed to Brightness2 project and the current work. The conceptual design was done by EV, LvE, HR; the computer code design and execution was done by EV, LvE and the interpretation and writing was done by all authors.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-88099-w>.

Correspondence and requests for materials should be addressed to E.V.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025