Marine Pollution:

Global issue, clustered research?

**Brie Sherow**

**Z5282148**

***Marine Pollution Bulletin***

**Abstract**

Why you did it 1-2: Marine pollution is a global issue that negatively impacts environmental, economic, and social systems. Discover if there are trends in topics and clustered research, identify gaps in geography and topics.

What you did 1-2: Extracted all articles within scopus that have keyword “marine pollution” and analysed them with bibliometrix package.

What you found 2-4:

What it means 1-2:

**Introduction**

*The Background*

Marine pollution is a global issue that has widespread negative impacts on ecological, social, and economic systems (Beaumont, et al. 2019). It is estimated that up to 12 million metric tonnes of plastic alone enters the ocean each year, costing roughly $13 billion in economic costs and impacting more than 800 marine and coastal species (United Nations 2020)***.*** The definition of marine pollution has evolved from a focus on the effects of industrial activity to a broader notion of the interdependence between human activity and nature (Tomczak 1984). The first widely accepted definition resulted from the United Nations *Convention of the Law and the Sea* in 1982. Marine pollution was defined as the ‘introduction by man, directly or indirectly, of substances or energy into the marine environment, including estuaries, which results or is likely to result in such deleterious effects as harm to living resources and marine life, hazards to human health, hindrance to marine activities, including fishing and other legitimate uses of the sea, impairment of quality for use of sea water and reduction of amenities’ (United Nations 1982).

*The Problem*

The environmental, economic, and health implications of marine pollution are global in scale; however the research is not standardised and the full impacts are not understood (Galgani 2015). The negative impact of marine pollution is widely accepted by scientists, policy makers, and the general public although knowledge gaps make it difficult to address the issues (Bonanno and Orlando-Bonaca 2018).

*The Aims (rename to be less repetitive)*

This paper aims to reveal clusters in geographies and topic areas of published marine pollution research in four parts. The first is a high level analysis of the overall literature, including annual publications, productive authors, and highly cited papers. The second is a geographic analysis of high producing countries. The third is a network map of intra and international collaborations based on author addresses. The fourth is a visualisation of author keywords in order to determine evolution of topic areas. Determining geographic concentrations, patterns of collaboration in research, and the evolution of topic areas will help us see how our understanding of marine pollution and its impacts have changed over time and where there may yet be gaps.

**Materials and Methods**

*Research Methodology*

I analysed the scientific literature using a bibliometrics study, a method of assessing published records used to evaluate research trends. Bibliometrics uses information on authors, citations, and keywords in order to describe the structure and evolution of publications throughout time (Nakagawa et al 2019).

*Data Collection*

In order to analyse marine pollution literature I extracted article metadata from SCOPUS database and created the dataset used in this analysis. In the first instance, I queried SCOPUS using the following string search:

EXACTKEYWORD ( "marine pollut\*" )  AND  ( LIMIT-TO ( SUBJAREA ,  "ENVI" ) )  AND  ( LIMIT-TO ( DOCTYPE ,  "ar" )  OR  LIMIT-TO ( DOCTYPE ,  "cp" )  OR  LIMIT-TO ( DOCTYPE ,  "re" ) )  AND  ( LIMIT-TO ( LANGUAGE ,  "English" ) )

This instructs Scopus to search for records tagged with the keyword ‘Marine Pollution’ and limited to the subject area of ‘Environmental Science’. I used a filtering process to include only publications in English and to limit document types to articles, conference papers, and reviews. I excluded articles published in 2020 in order to assess only full calendar years.

Once I had downloaded the records with full citation rmetadata, I then used *bibliometrix* package (Aria and Cuccurullo 2017) in R (R Core Team 2020) to transform the Scopus BibTeX file into a dataframe used for analysis.

*Descriptive Analysis*

I used core bibliometrix outputs to determine the total number of articles published each year, the country where the research is based, and top authors assessed by number of articles published and total number of citations.

*In-depth Analysis*

I used *tmap* package (Tennekes 2018) in R in order to visualise the spatial distribution of marine pollution publications. To further assess the geography of marine pollution research I used network mapping with *igraph* package (Csardi and Nepusz 2006) in R to visualize intra and international collaboration networks. I analysed this network using the Fruchterman-Reingold layout to reflect the structure of the networks while distributing connected vertices near each other without being drawn so close that they are obscured. (Fruchterman & Reingold 1991). I then used a different layout in order to visualise the individual vertices clearly. In order to assess the changes in topic areas over time I extracted all author keywords and calculated the frequency of each word. I then created a wordcloud for early research (1970-2009) and another for recent research (2010-2019) using *wordcloud2* package (Lang and Chien, 2018). Supplementary material for this paper may be found at [GitHub](https://github.com/brieaspasia/mp-diagnostics).

**Results**

*Descriptive Analyses*

*Publications*

After searching and filtering records I identified a total of 9,757 publications on marine pollution between 1970 and 2019. The main subject disciplines were Environmental Science (9,757 articles), Earth and Planetary Sciences (4,294 articles), Agricultural and Biological Sciences (4,261 articles), Chemistry (975 articles), and Pharmacology, Toxicology and Pharmaceuticals (615 articles).

The top journals were Marine Pollution Bulletin (3,179 articles), Science of the Total Environment (563 articles), Environmental Science and Technology (445 articles), Environmental Pollution (433 articles), and Chemosphere (385 articles).

A close up of a map

Description automatically generated

Figure 1 Articles published per year

The first published article to record the keyword ‘marine pollution’ was in 1970 but it wasn’t until 1985 that the number of publications per year passed ten. The number of publications per year remained marginal until 1995 marked the year that the number rose above 100. The annual percentage growth rate from 1970-2019 is 15% and much of that has been in the last ten years.

A close up of a map

Description automatically generated

Figure 2 Influential papers

I have identified six influential papers with over 100 citations each. All are about plastics; three about microplastics, two about plastic waste in general, and one about the impact of plastics on marine life. Four out the six are reviews, and five out of the six were published within the last ten years.

*Authors*

There are 43,788 author appearances in the literature representing 26,790 unique authors. A graph of top producing authors largely follows the trends in annual production, with authors increasing amount of publications throughout time rather than being staggered throughout time. Many of the top producing authors began publishing around 2000 and have either been publishing steadily or increasing their publication rate since then.

A screen shot of a computer

Description automatically generated

Figure 3 Top producing authors

A table of the top twenty cited authors has little overlap with the top producing authors, demonstrating that the most influential authors are not publishing most frequently. Only four authors appear in both top cited and top producing; S. Tanabe, J. Wang, Z. Wang, and WX Wang. They are affiliated with institutions in Japan, China, and Hong Kong; and are largely researching micropastics and heavy metals.

I visualized relationships between the top authors using a chord diagram (Gu 2014) to represent the bibliographic coupling network with authors as nodes and shared references as edges. This can aid in identifying research structure even in very recent articles that have not had time to be cited prolifically (Belter and Seidel 2013). In this analysis we see that the citation structure of the literature as a whole is quite equally spread, with the top authors all referencing similar material.

**A picture containing device

Description automatically generated**

Figure 4 Author's bibliographic coupling

*Geography*

I used *bibliometrix* package in R to extract the author addresses in order to map their affiliation. I then used *tmap* package in R to reveal the spatial trends in marine pollution publishing. The research is widespread, with 102 countries total whose institutions have published about marine pollution. Of those, more than 50% are made up of the top 7 so it is also quite clustered. The US and China are key players and completing the top 50% are Italy, Spain, the UK, Canada, and Australia.

Other notable countries are Japan, Brazil, and France. The research institutions are based predominantly in the northern hemisphere and in developed nations.

A close up of a map

Description automatically generated

Figure 5 Researcher affiliations

*Intra and International Collaborations*

I used *igraph* to create a network plot of the top 75 research institutions and 747 edges of collaboration between them. I used Fruchterman-Reingold layout as a means to display the collaboration groupings as it effectively distributes vertices based on collaboration edge link (Fruchterman & Reingold 1991). Chinese institutions are tightly clustered intra-nationally. Other Asian nations collaborate internationally, dominated by Japan, Hong Kong, Korea, and India. Portugal collaborates intra-nationally with strong ties between three institutions. The UK is loosely tied with continental Europe but collaborating primarily intra-nationally while other European nations collaborate primarily internationally. Canadian universities are clustered intra-nationally but also have ties with Arctic specific institutions internationally. Brazil is grouped loosely with USA rather than with other southern hemisphere institutions.  Australian universities are spread between clusters based in the UK, USA, and Canada.

A close up of a map

Description automatically generated

Figure 6 Intra and International Collaborations

**A close up of a map

Description automatically generated**

Figure 7 Institutions and their Collaborators

*Research Topics*

I used *bibliometrix* to determine extract 16,770 total author keywords recorded. I separated the data into a subset of early research (1970-2009) and a subset of recent research (2010-2019) and visualised the topics by analysing frequency of author keywords.

The top five early keywords are related to heavy metals, sediments, and bioaccumulation. Other notable keywords in the top fifty include those related to metals (metals, trace metals, metal, mercury, lead, cadmium, copper), and other pollutants such as nutrients, pesticides, DDT, organochlorines, sewage, oil spills, TBT, tributyltin (an anti-fouling agent). There is evidence of impact assessment with words such as bioaccumulation, monitoring, water quality, biomarkers, toxicity, risk assessment. Other notable include PAH (from emissions) and PCBs (from plastics, electronics, and paints). The only locations listed are the Baltic Sea, the Mediterranean Sea, and both Antarctica and the Arctic rounding out the top fifty words.

**A close up of a logo

Description automatically generated**

Figure 8 Early Research Topics (1970-2009)

The subset of recent publications (2010-2019) also mentioned heavy metals and sediments in the top five, however they were displaced from their top spot by marine debris, microplastics, and oil spills. Within the top fifty there were many notable overlaps with the early research although there was less mention of specific heavy metals. In addition, there were many new plastics related keywords including marine litter, microplastics, plastic pollution, plastics, and plastic debris. Rounding out the top fifty were bioremediation, climate change, and pharmaceuticals. Locations mentioned include Mediterranean Sea, Gulf of Mexico, the Arctic, and the Baltic Sea.

**A close up of a logo

Description automatically generated**

Figure Recent Research Topics (2010-2019)

**Discussion, limitations, and future directions**

*Author networks*

It’s clear that marine pollution research is growing and that topic areas are increasing and diversifying. However initial findings often led to further questions making it difficult to draw conclusions about the gaps in knowledge.

The quantity of knowledge produced does not necessarily correspond to its influence. We can see that the authors producing the most research are not the same as the authors that are the most heavily cited. From observation, the highest producing authors are largely writing about heavy metals and organic pollutants. This makes sense as these topic areas have been present in the research for a longer period of time. The most highly cited papers, however, are all about plastics and debris. This may speak to new trends and recent growth in the research that is heavily focused on plastics. In order to more fully answer these questions it would be interesting to analyse the focus areas of these authors using *rscopus* package in R in order to determine if and where there are overlaps between high producing and highly cited authors. Additionally, for the sake of this analysis I measured influential papers by the number of citations, but one could argue that there are other, more relevant ways to measure influence. It could be interesting to take a topic area and determine a means to assess whether research has led to policy changes or public awareness and engagement. Involvement by government agencies or the general public may be a more accurate means of assessing influence, even if it would be more difficult to assess than citation records.

The bibliographic coupling analysis demonstrates that the research is quite insular, and that authors seem to be citing the same sources. This visualisation isn’t particularly helpful in determining the structure of the research, my sense is that a different visualisation such as a dendrogram would be able to show a history of the citation network. With a dendrogram we’d be able to identify influential papers and the successive research that these papers inspired.

*Geography*

Pollution does not respect national boundaries, yet research institutions do. Mapping is effective for visualising the global extent of the literature; however the location of the research institution is not necessarily the location of the field work. The map in Fig 5 does not display many countries in the developing nations of the tropics who often have the least policy surrounding pollution and are the most impacted by its negative impacts. In order to determine if the research site corresponds to the author’s national affiliation we would need to use text mining to assess site locations in studies more specifically. I believe with this additional analysis we would find many research sites in the tropics.

Additionally, the geographic analysis produced in *bibliometrix* is not adjusted for research effort, meaning that nations with fewer total institutions have less opportunity for publishing even if they are globally significant in the research despite their size. Spatial maps also cause visual bias towards geographically large nations that are easier to notice. For example, Hong Kong is the 12th most producing nation, but there is no representation of it in the spatial file.

The network analysis clearly showed intra and international collaborations, however the analysis would be more meaningful if we had more information about the nature of these collaborations. We get some clues from the names of the research institutions, for example the Canadian Universities have strong ties to Arctic specific institutions intenationally. When compared to the keyword analysis, the prominence of both the Mediterranean and the Baltic Seas suggest that international collaborations in Europe reflect the nations’ shared coastlines. The lack of prominent locational keywords in Asia or the Americas may be due to the nature of the research being more focused on the nature of pollution rather than the ecosystem health. These are only guesses until we quantify the linkages between keywords and collaboration clusters.

*Topic areas*

Wordclouds are visualisation tools that can give clues about where we might follow up in a more rigorous analysis.  In order to fully understand the evolution of the literature we would need to quantify the keywords. There are many duplicate records that skew the results, for example we could consolidate metal, heavy metal and trace metal as well as the plurals of those words. Depending on the questions we were asking we could include the individual names of each type of metal in that group. However, some duplicate records would be more subjective in their groupings. For example; biomarkers, bioaccumulation, and biomonitoring could refer to similar topics in some instances but not in others.

Quantifying the changes in keyword frequency would help us understand whether each topic area is growing or shrinking in terms of research production and citation influence. We may find that the growth in plastics research is starting to overshadow the prominence of earlier research topics, or we may find that all topic areas are growing relatively equally.

It could be interesting to map the history of citizen science in the literature as public engagement is an important indicator of the global awareness of the issues. Other new issues in the recent keywords include climate change and pharmaceuticals. Both of these issues speak to our evolving understanding of the connections between the ocean and many aspects of our every day lives.

**Broader Implications**

The negative impacts of marine pollution are widely accepted by scientists, policy makers, and the general public. These negative impacts effect public health as well as ecological, economic, and social systems. It is important to understand how this research has shifted and whether there are regions, habitats, and topics that remain marginal. The positive influence of research occurs when scientists are able to communicate their findings with policy makers and the general public. In this study I’ve made progress towards identifying the clusters in the geography and topic areas of marine pollution. Further study could illuminate gaps in the literature that could inform future research and direct resources, particularly towards marginalised regions that may be most impacted by marine pollution while having the least resources to manage it. Measuring links between research and corresponding policy could illuminate the influence of academic findings on certain topic areas. If current trends continue, the literature on marine pollution will continue to grow as we explore more of the ocean and discover all of the ways that our lives rely on it.

*Acknowledgements*

Thanks to Shinichi Nakagawa, Losia Nakagawa-Lagisz for their support in the research and coding aspects of this project. Thanks to Leigh Marine Lab providing a productive study space and field work opportunities that gave temporary respites from R.

# Bibliography

Aria, M., and C. Cuccurullo. 2017. "bibliometrix: An R-tool for comprehensive science mapping analysis." *Journal of Informetrics* (Elsevier) 959-975.

Beaumont, N., M. Aanesen, M. Austen, T. Börger, M. Cole, T. Hooper, P. Lindeque, C. Pascoe, J. Clark, and K. Wyles. 2019. "Global, ecological, social, and economic impacts of marine plastic." *Marine Pollution Bulletin* 189-195.

Belter, C., and D. Seidel. 2013. "A bibliometric analysis of climate engineering research." *WIREs Clim Change* (Wiley and Sons) 4: 417-427.

Bonanno, G., and M. Orlando-Bonaca. 2018. "Ten inconvenient questions about plastics in the sea." *Environmental Science and Policy* 146-154.

Csardi, G., and T. Nepusz. 2006. "The igraph software package for complex network research." *http://igraph.org.* InterJournal, Complex Systems 1695. http://igraph.org.

Environment Program, United Nations. 2020. *United Nations Environment Program - Oceans and Seas.* Accessed June 2020. https://www.unenvironment.org/explore-topics/oceans-seas/what-we-do/working-regional-seas/marine-litter.

Fruchterman, T., and E. Reingold. 1991. "Graph Drawing by Force-directed Placement." *Software -- Practice and Experience* 1129-1164.

Galgani, F. 2015. "Marine litter, future prospects for research." *Frontiers in Marine Science* 1-5.

Ghayebzadeh, M., H. Taghipour, and H. Aslani. 2020. "Estimation of plastic waste inputs from land into the Persian Gulf and the Gulf of Oman: An environmental disaster, scientific and social concerns." *Science of the Total Environment.*

Gu, Z. 2014. "circlize implements and enhances circular visualisation in R." *Bioinformatics.*

Lang, D., and G. Chien. 2018. *wordcloud2: Create Word Cloud by 'htmlwidget'.* January 03. Accessed July 2020. https://cran.r-project.org/web/packages/wordcloud2/.

Nakagawa, S., G. Samarasinghe, N. Haddaway, M. Westgate, R. O'Dea, D. Noble, and M. Lagisz. 2019. "Research Weaving: Visualising the Future of Research Synthesis." *Trends in Ecology and Evolution* 34 (3): 224-238.

R Core Team. 2020. "R: A language and environment for statistical computing." *http://www.R-project.org/.* R Foundation for Statistical Computing.

Syed, S., L. Aodha, C. Scougal, and M. Spruit. 2019. "Mapping the global network of fisheries science collaboration." *Fish and Fisheries* 830-856.

Tennekes, M. 2018. "tmap; Thematic Maps in R." (Journal of Statistical Software) 1-39.

Tomczak, M. 1984. "Defining Marine Pollution: A comparison of definitions used by international conventions." *Marine Policy* (Butterworth & Co Ltd) 311-322.

United Nations. 2020. *UN Ocean Conference.* Accessed June 2020. https://www.un.org/en/conferences/ocean2020/facts-figures.

United Nations, Office of Legal Affairs. 1982. *United Nations Convention of the Law and the Sea.* December 10. Accessed June 2020. https://www.un.org/Depts/los/convention\_agreements/texts/unclos/part1.htm.

Wickham, H. 2016. "ggplot2: Elegant Graphics for Data Analysis." New York: Springer-Verlag.

Wright, H. 1962. "Oil and troubled waters." *Drilling and Production Practice* 90-98.