**Double spaced, A4, 12pt font**

**Title**

Marine Pollution: Global issue, clustered research?

**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*

This paper aims to map the clusters in geographies and topics in published marine pollution research in order to determine how it has changed over time, where it is concentrated geographically and where there are still gaps. A first look at the literature indicates that early articles focus on oil spills and industrial pollution (Wright 1962) while much of the current research centres around debris, specifically plastics. Marine pollution does not respect national boundaries, yet research clusters do, and may not represent the regions most impacted by marine pollution. Thus, I will record where published research of marine pollution comes from and how the topic areas have changed over time.

**Materials and Methods**

*Data Collection*

Article metadata was extracted from SCOPUS in order to produce the data 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’. A filtering process was used to include only publications in English and to limit document types to articles, conference papers, and reviews. Articles published in 2020 were excluded in order to assess only full calendar years. 

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

Top journals are Marine Pollution Bulletin (3,179 articles), Science of the Total Envronment (563 articles), Environmental Science and Technology (445 articles), Environmental Pollution (433 articles), and Chemosphere (385 articles).

I downloaded these records and then used *bibliometrix* package (Aria and Cuccurullo 2017) in R (R Core Team 2020) to transform a 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) In order to assess the changes in author keywords over time I created a wordcloud for early research (*years*) and another for recent research (*years*) using *wordcloud2* package in R.

Code and data availability – github link?

**Results**

*Descriptive analyses*

* Articles published per year, with discussion of top 5 cited.

A close up of a map

Description automatically generated

Figure

* Top producing authors not staggered, all increasing with time, largely not correlated to the top cited authors.

A screen shot of a computer

Description automatically generated

Figure

*Geography*

* A few countries represent the majority of the research despite the issues and solutions necessitating global impact.

A close up of a map

Description automatically generated

Figure

* We have record of where the research institution is based, but this is not necessarily indicative of where the research is taking place.
* Only 5 countries produce more than 5% of the research each, totalling over 40% total.
* Dominated by the USA and China. The UK publishes a lot, and there’s a cluster in the Mediterranean dominated by Italy and Spain.
* Other notable countries in the top ten are Canada, Aus, Japan, Brazil, and France. Relatively spread out around the world.
* Many sources of pollution and those impacted most by it are not necessarily represented in the research. Pollution does not respect national boundaries, research and solutions must be global in scale, regions must not remain marginal.

*Intra and International Collaborations*

The initial network of affiliations contains 75 research institutions and 747 edges. Fruchterman-Reingold layout for clustering. A close up of a map

Description automatically generated

Figure

China is tightly clustered intra-nationally, as is Portugal with fewer institutions.  Northern Europe is very collaborative internationally, Canadian universities are clustered intra-nationally but also have ties with Arctic specific institutions internationally. UK is loosely tied with continental Europe, but collaborating primarily intra-nationally.  Brazil is grouped loosely with USA rather than with other southern hemisphere institutions.  Japan collaborates with Hong Kong, Korea, and India. Australia Universities are grouped with USA, UK, and Canada - are they not working with each other?

A close up of a map

Description automatically generated

Figure

*Research topics*

Wordclouds are great for visualization, and here we can get some clues about where we might follow up in a more rigorous analysis.  Early research shows a focus on heavy metals, and oil spills and is concerned with risk. Bioaccumulation and eutrophication suggests further study into agricultural and industrial pollution.  There isn’t much focus on plastics, bar mention of PCBs and chlorinated compounds.

A screenshot of a cell phone

Description automatically generated

Figure

Recent research still shows metals, pesticides, oil spills and eutrophication, but now we see many more plastics related words and a focus on debris and litter.  There is mention of climate change and PAH (related to emissions). The appearance of citizen science suggests more engagement and a focus on social implications of pollution.  Also interesting to see the focus on pharmaceuticals.

**A picture containing food

Description automatically generated**

Figure

**Discussion**

*Key Findings*

It’s clear that marine pollution research is growing and topic areas are growing and diversifying.

*Status of hypothesis*

Further analysis is needed to quantify how that is happening and to draw conclusions about gaps in research. Further study could analyse keywords by research cluster to determine the gaps in research. We have record of where the research institution is based, but this is not necessarily indicative of where the research is taking place.

*What went wrong and why*

* Initial findings were often unclear and further research is needed before conclusions can be made.
* Bibliometrix has some beautiful preset graphics, however they make it very difficult to extract the information in order to explore, modify, and customise their outputs.
* The most productive authors are not represented in the most cited papers. Is this due to plastics research growing more than other topics? Could draw more conclusions from author data with thorough bibliographic coupling analysis and rscopus package to
* Wordclouds are effective for visualising data, but we must quantify the changes in keywords over time in order to fully understand how the research is evolving.
* More thorough text mining is needed in order to more fully understand the topics covered and the gaps in knowledge.

*Broader Implications*

* How has research changed over time? Quantify keywords. Initial analysis confirms that topic keywords are diversifying and expanding to broader subject areas.
* Gaps in knowledge. Methods and methodologies make it difficult to compare. Top cited papers primarily represent plastic, are other topic areas being overlooked?
* The research does appear to be spread globally, however are certain subject areas more common in certain research clusters?
* How has this research influenced policy? Measure links between university and govt research institutions? Compare to policy changes in topic areas?

*Acknowledgements*

Thanks to Shinichi, Losia, and Leigh Marine Lab for a study space.