Postprint: Analysis of Research Hotspots and Trends in Non-point Source Pollution Models Based on CNKI and WOS

Authors: Jing Yande, Fan Rui

Date: 2018-06-09T00:00:00+00:00

Abstract

Using journal papers from 1986-2016 in the China National Knowledge Infrastructure (CNKI) Academic Journal Network Publishing Database and WEB OF SCIENCE (WOS) as data sources, literature was retrieved using non-point pollution, Diffused pollution, or non-point source pollution and Model as search themes, yielding a total of 1,474 domestic and 1,046 foreign publications. Co-word analysis methods were employed in combination with CiteSpace, SPSS, and Bibexcel software to analyze the retrieval results, and Ucinet and NetDraw software were used to generate co-word network visualizations. Based on the processed results, a systematic review and analysis of research trends in non-point source pollution models was conducted, examining the current status and development of non-point source pollution model research from multiple perspectives, including keywords, temporal distribution, publishing journals, authors, highly cited articles, etc. On this basis, the research overview and hotspots in the field of non-point source pollution models both domestically and internationally were analyzed, clarifying the exploratory directions for non-point source pollution model research and providing a theoretical basis for future studies in this area.

Full Text

Preamble

Acta Ecologica Sinica (ChinaXiv Partner Journal)
Vol. 38, No. 11, June 2018

DOI: 10.5846/stxb201706021019

Research Hotspots and Trends Analysis of Non-Point Source Pollution Models Based on CNKI and WOS

Key Laboratory of Wetland Ecology and Environmental Protection of Nansi Lake, Universities in Shandong Province
School of Geography and Tourism, Qufu Normal University

Abstract

This study utilizes journal papers from the China National Knowledge Infrastructure (CNKI) and Web of Science (WOS) databases as data sources. Using search terms including "non-point pollution," "diffuse pollution," "non-point source pollution," and "model," we employed co-word analysis methods and software tools including CiteSpace, SPSS, Bibexcel, Ucinet, and NetDraw to analyze the retrieval results and map co-word network visualizations. Based on the processed results, we systematically reviewed and analyzed research trends in non-point source pollution models.

Non-point source pollution, also known as diffuse pollution, represents a major societal challenge, with agricultural non-point source pollution contributing the largest share. Excessive nitrogen and phosphorus emissions from agricultural non-point sources are direct causes of water eutrophication. Non-point source pollution is characterized by complex formation processes, wide distribution, complex influencing factors, long incubation periods, and significant hazards, making monitoring and management difficult. Due to the difficulty in calculating non-point source pollution loads, quantitative research on non-point source pollution is particularly important, with modeling serving as a crucial methodology.

Co-word analysis, derived from citation coupling and co-citation analysis in bibliometrics, is widely used to reveal research themes and hotspots in a field, as well as to explore the historical evolution and development trends of research topics. Watershed-scale non-point source pollution mechanism models help managers quantify pollution loads and rapidly identify critical source areas at the spatial scale, becoming one of the most direct and effective approaches for studying non-point source pollution.

Although research on non-point source pollution models in China has gradually increased, few papers have used bibliometric methods to comprehensively understand the research status and development process of non-point source pollution models both domestically and internationally. This paper addresses this gap by using literature from CNKI and WOS, applying co-word analysis combined with relevant software to conduct a macro-level analysis of research hotspots and development directions in non-point source pollution models, aiming to provide direction for future research in this field.

Keywords: non-point source pollution, SWAT model, GIS, co-word analysis, research hotspots

1. Data Sources

This study uses the CNKI knowledge discovery network platform and the Web of Science Core Collection as retrieval sources. Both databases contain extensive collections of high-quality research papers and represent reliable data sources for bibliometric analysis, forming the foundation of this study. Combined with co-word analysis methods, we analyzed keywords from published articles and highly cited papers to present the research overview, hotspots, and trends in the field of non-point source pollution models using software including CiteSpace, SPSS, Ucinet, and NetDraw.

Foreign Literature Analysis

1.1 Temporal Distribution of Foreign Publications on Non-Point Source Pollution Models

Using the Web of Science Core Collection as the data source, we searched for "Model" combined with "non-point pollution OR diffused pollution OR non-point source pollution" with a specified time range. The search yielded a total number of documents, with annual publication quantities shown in [Figure 1: see original paper]. The number of foreign articles on non-point source pollution models reached two small peaks in 2014 and 2016, with publication numbers indicating that as water environmental pollution becomes increasingly severe, the damage caused by non-point source pollution to water environments has received high attention, becoming a hot research topic.

Foreign research on non-point source pollution models has consistently maintained a high level, with occasional declines during the process. However, overall, annual publication numbers have remained above a certain threshold, showing an almost linear growth trend. Before a certain year, the count was zero. [Figure 1: see original paper] shows the distribution of foreign non-point source pollution model literature, demonstrating that academic attention to non-point source pollution models remains relatively high.

1.2 Keyword Frequency Analysis

Using the bibliometric analysis software Bibexcel, we counted keyword frequencies. Since identical keywords may be expressed differently, we merged synonymous keywords. After removing the most frequent keyword "non-point source pollution," the remaining keywords reflect international research directions and hotspots in non-point source pollution to a certain extent. We selected the top high-frequency keywords for analysis.

Research hotspots appear in "water quality," "non-point source," and "SWAT," with keyword frequencies exceeding certain thresholds, accounting for 30.96% of all statistical keywords—far above the average level. This indicates relatively specific research scope. SWAT model research accounts for 8.2% of statistical keywords, significantly higher than other models, making it the hottest research model in non-point source pollution modeling due to its accurate and powerful simulation capabilities and user-friendly features.

Keywords related to "Phosphorus," "Nitrate," and "Nitrogen" also show relatively high frequencies, proving that non-point source pollution is closely related to chemical elements, especially nutrients. This aligns with Guo Linlin' s research [14]. In the statistical keywords, nutrients such as nitrogen and phosphorus are important elements causing water environmental pollution and are primary causes of surface water eutrophication, remaining important research topics for solving water quality issues.

Keywords related to agriculture and land use, such as "Agriculture" and "Land use," show relatively high average proportions, indicating that agricultural production contributes significantly to non-point source pollution. Since agricultural non-point source pollution has the largest contribution rate [15-16], research on land use patterns and rational urban-rural layout will be key focus areas in future non-point source pollution research. Mathematical models are important tools for quantitatively describing and evaluating non-point source pollution, and GIS technology can be integrated with mathematical models to analyze the temporal and spatial distribution characteristics of non-point source pollution, serving as an important tool in non-point source pollution research.

1.3 Keyword Visualization Analysis

Although the above keywords reflect research hotspots and directions in non-point source pollution models to some extent, frequency alone cannot reveal relationships between them. We paired high-frequency keywords and counted their co-occurrence. To reduce the impact of keyword association on visualization results, we used the Ochiai coefficient statistical index [17] with the following formula:

Ochiai coefficient = (Co-occurrence frequency of words A and B) / √(Frequency of word A × Frequency of word B)

Using Bibexcel software, we constructed a 30$×$30 co-word matrix and used NetDraw and Ucinet software to generate keyword co-occurrence network visualizations. The network shows interwoven relationships among 30 high-frequency keywords, where node size represents keyword centrality (larger nodes indicate higher centrality), and connecting lines represent co-occurrence frequency between keyword pairs (thicker lines indicate higher frequency and closer relationships) [13].

This interwoven network centers around high-frequency keywords such as "water quality," "Nitrogen," "BMPs," "Nitrate," "SWAT," and "Phosphorus," forming the core structure. This indicates that these are hot research directions in non-point source pollution modeling, consistent with keyword frequency statistics. It also demonstrates international emphasis on non-point source pollution issues and model application in exploring solutions. Nutrients like phosphorus and nitrates remain important research topics, as eutrophication has become a significant international concern, with researchers exploring best management practices (BMPs).

[Figure 2: see original paper] shows the foreign keyword co-occurrence network visualization.

1.4 Citation Frequency Analysis

Paper citation rate refers to the number of times a scientific paper is cited by other literature, serving as a measure of a country' s research recognition by other nations or institutions. We selected the top 15 papers by average annual citation frequency, compiling their publication journals, total citation frequency, average annual citation frequency, and other metrics.

The paper with the highest average annual citation frequency is by Ongley, Edwin D, published in *ENVIRONMENTAL POLLUTION*, which studied pollution assessment in China. This paper summarizes methods for evaluating non-point source pollution in China and research conditions for agricultural non-point source pollution, with important reference value for studying pollution loads of key chemical elements in different Chinese watersheds. Its total citation frequency is also relatively high.

JOURNAL OF HYDROLOGY published the most papers among the top 15, representing a core force in international non-point source pollution research, accounting for 26.67% of the statistical years. Research on non-point source pollution models abroad had already reached relatively advanced levels before a certain year.

shows the publication years and average annual citation frequencies of foreign papers.

1.5 Keyword Clustering Analysis

Clustering analysis is a scientific and effective method for studying relationships. When clustering high-frequency keywords, we used hierarchical clustering, treating each keyword as a separate class initially, then merging the closest two classes repeatedly until all keywords were grouped into one class [18]. Based on the 30$×$30 dissimilarity matrix constructed above, we performed clustering analysis using software to generate a clustering dendrogram [Figure 3: see original paper].

Combined with relationships among non-point source pollution models, the 30 high-frequency keywords can be divided into several main clusters, with keywords in each cluster shown in . The analysis reveals that: (1) one cluster focuses on point and non-point source water pollution, especially severe eutrophication; (2) another cluster covers water pollution models and framework directives, highlighting the seriousness of water pollution caused by non-point sources and scholars' attention to this issue; (3) a more complex cluster involves various non-point source pollution models and tools for analyzing their temporal and spatial distribution characteristics.

To more clearly visualize keyword clusters, we used multidimensional scaling analysis in statistical software to construct a multidimensional scaling visualization map. The distribution of keywords from the same cluster in this map is basically consistent with hierarchical clustering results, showing classification patterns more intuitively.

shows the cluster division of foreign "non-point source pollution models."

Domestic Literature Analysis

2.1 Temporal Distribution of Domestic Publications on Non-Point Source Pollution Models

Using the CNKI academic journal network publication database as the data source, we searched for "non-point source pollution OR diffuse pollution" AND "model" within a specified range. The search yielded a total number of papers, with annual publication quantities shown in [Figure 5: see original paper].

The figure shows that: (1) although some scholars studied non-point source pollution earlier, it did not attract much attention in China at that time, with relatively few related studies and very few published papers; (2) research papers on non-point source pollution models in China began to increase, showing an upward trend; (3) the number peaked in a certain year, with related literature reaching a certain quantity. As pollution problems become increasingly severe and the country attaches greater importance to sustainable development, the number of papers published annually has remained above a certain level. Although the number declined in some years, overall, research on non-point source pollution in China has gradually deepened.

2.2 Keyword Frequency Analysis

Using the bibliometric analysis software CiteSpace, we counted keyword frequencies. Since keywords with the same meaning may be expressed differently in various articles, we merged synonymous keywords and removed the highest-frequency keyword "non-point source pollution." After consolidation, we selected high-frequency keywords for statistical analysis to examine their interrelationships. These keywords reflect research hotspots and directions in China' s non-point source pollution model field to a certain extent.

Selected keywords and their frequencies are shown in . "Agricultural non-point source pollution" accounts for 11.88% of statistical keywords, second only to "non-point source pollution," indicating that China focuses on agricultural non-point source pollution issues. The SWAT model is the most commonly used model in non-point source pollution research, representing a domestic research hotspot that corresponds to international SWAT model research. AnnAGNPS also shows relatively high frequency, indicating these models are important tools for studying non-point source pollution in China.

Among high-frequency keywords, many relate to agricultural production and land, such as "agricultural non-point source pollution," showing that agricultural production is an important factor causing non-point source pollution. Pesticides and fertilizers used in agricultural production flow into rivers, causing serious pollution, making it a critical issue for China to address.

The keyword "Geographic Information System" shows relatively high frequency, accounting for 7.62% of statistical keywords—far above the average. GIS has been widely applied in the non-point source pollution field due to its capabilities in collecting and analyzing geographic information. This corresponds to foreign keyword statistics, indicating GIS is an important element in non-point source pollution research both domestically and internationally.

2.3 Keyword Visualization Analysis

The domestic high-frequency keyword co-occurrence network visualization [Figure 6: see original paper] shows that, apart from "non-point source pollution," keywords such as "SWAT," "GIS," and "agricultural non-point source pollution" occupy central positions. Therefore, they represent research hotspots and directions in China' s non-point source pollution models, consistent with conclusions from high-frequency keyword statistics. This indicates that using models to solve rational land use allocation, address agricultural non-point source pollution, and manage nitrogen and phosphorus elements are currently important research issues and areas in non-point source pollution.

2.4 Citation Frequency Analysis

We statistically analyzed the top 15 domestic papers by average annual citation frequency. The paper with the highest average annual citation frequency, published in *Agricultural Technology Economics*, used statistical analysis and econometric modeling to study the impact of chemical fertilizers on agricultural non-point source pollution. The study distributed questionnaires to understand farmers' basic characteristics more authentically, then used regression analysis to study factors influencing fertilizer application, providing reference value for agricultural non-point source pollution research.

Agricultural Technology Economics and *Shanghai Environmental Science* are two important journals publishing such papers, representing core forces in domestic non-point source pollution research. This corresponds to statistics on publication years and numbers. Domestic research on non-point source pollution models mainly began after a certain year, with significantly increased publications.

shows the total citation frequency and average annual citation frequency of domestic papers.

2.5 Keyword Clustering Analysis

Using the same method as the foreign non-point source pollution model research described above, we obtained a hierarchical clustering dendrogram [Figure 7: see original paper] and multidimensional scaling map. The cluster division table shows that domestic clusters mainly involve: (1) nitrogen and phosphorus nutrients, corresponding to a major research direction in foreign non-point source pollution; (2) tools for solving non-point source pollution, such as GIS, which can more accurately describe the spatial distribution of non-point source pollution; (3) non-point source pollution models, with analysis focusing on the Taihu Lake basin and Three Gorges Reservoir area.

Conclusions and Discussion

From a quantitative perspective, the number of papers published both domestically and internationally increased after a certain year, indicating that research on non-point source pollution models has gained attention worldwide. During certain years, China published more papers than foreign countries, showing that domestic research on non-point source pollution models has received attention, corresponding to China' s proposed sustainable development and environmental protection initiatives.

From a research hotspot perspective, foreign research hotspots include "water quality," "Nitrogen," "Watershed," and "Phosphorus," while domestic hotspots focus on "SWAT," "GIS," and agricultural non-point source pollution. Domestic and foreign research hotspots share similarities, but China places more emphasis on GIS research and application and the contribution of agricultural production to non-point source pollution, while foreign countries focus more on watershed-based non-point source pollution model exploration.

From a cluster classification perspective, domestic and foreign cluster divisions are basically similar, which can be summarized into four main aspects: non-point source pollution models, nutrients, GIS technology, and development processes.

This study selected English keywords with cumulative frequency reaching a certain threshold and Chinese keywords with cumulative frequency reaching another threshold as research hotspots. However, we do not exclude the possibility that some low-frequency keywords may become research hotspots in the future. The results are influenced by the completeness of keyword statistics, professional levels, and disciplinary categories. Using co-word analysis to explore disciplinary fields has certain limitations.

References

[1] Remote sensing extraction and quantitative assessment of non-point source risk in drinking water sources. *Environmental Science Research*, 2013, 26(12): 145-150.

[2] Estimation of agricultural non-point source pollution load in bay based on watershed units: A case study of Laizhou Bay. *Acta Ecologica Sinica*, 2014, 34(1): 173-181.

[3] Comprehensive evaluation of agricultural non-point source pollution in the Gulin River basin, Gulin County section. *Research of Soil and Water Conservation*, 2016, 23(5): 327-331.

[4] Research on non-point source pollution models. *Heilongjiang University Engineering Journal*, 2002, 23(3): 124-128.

[5] Comparison and selection of non-point source pollution load models. *Environmental Science & Technology*, 2011, 2(2): 49-53.

[6] Agricultural non-point source pollution load and current status evaluation: A case study of the Dasu River area. *Journal of Agro-Environment Science*, 2013, 36(5): 176-182.

[7] Agricultural non-point source pollution load and current status evaluation. *Journal of Agro-Environment Science*, 2010, 29(S1): 1349-1355.

[8] Research and application of rapid identification methods for critical areas of watershed non-point source pollution: A case study of the Haihe main stream watershed. *Anhui Agricultural Sciences*, 2017, 45(11): 50-54.

[9] Le Marc M, Courtial J P, Senkovska E D, Petard J P, Py Y. The dynamics of research in the psychology of work from 1973 to 1987—from the study of companies to the study of professions. *Scientometrics*, 1991, 21(1): 69-86.

[10] Research status of ecological restoration based on bibliometric analysis. *Acta Ecologica Sinica*, 2016, 36(8): 2442-2450.

[11] Co-word analysis method—Principles and characteristics of co-word cluster analysis. *Information Science*, 2008, 27(7): 118-120.

[12] Edwards A C, Withers P J A. Transport and delivery of suspended solids, nitrogen and phosphorus from various sources to freshwaters in the UK. *Journal of Hydrology*, 2008, 350(3/4): 144-153.

[13] Social network analysis: A management research method and perspective. *Science and Technology Management Research*, 2015, 33(4): 149-153.

[14] Guo Linlin. Research on non-point source pollution models. *Environmental Science Research*, 2013.

[15] Lacroix A, Beaudoin N, Makowski D. Agricultural water nonpoint pollution control under uncertainty and climate variability. *Ecological Economics*, 2005, 53(1): 115-127.

[16] Zhuang Y H, Thuminh N, Niu B B, Shao E, Hong S. Research trends in non-point source during 1975—2010. *Physics Procedia*, 2012, 33(1): 138-143.

[17] Ochiai coefficient application in co-word analysis. *Information Science*, 2010.

[18] Hierarchical clustering method principles. *Journal of Information Science*, 2010, 30(24): 216-219.

[19] Comparison study on salt tolerance of different wheat varieties. *Journal of Xinyang Normal University (Natural Science Edition)*, 2007, 20(4): 442-444, 451-451.

Note: Figure translations are in progress. See original paper for figures.