Maritime Technology and Society
Vol. 4(3): 139-152, October 2025
https://doi.org/10.62012/mp.vi.44925

Review Article

Multivariate Risk Analysis of Echotoxic Chemicals of Ballast Water Chemicals Based on PCA and DSS Using ECOTOX GISIS Data

Ariyono Setiawan[1,*], Upik Widyaningsih[1], Anjar Pamungkas[1], Abdul Razak Bin Abdul Hadi[2], Deshinta Arrova Dewi[3]

[1*]Politeknik Pelayaran Surabaya, Indonesia
[2] University of Kuala Lumpur, Malaysia
[3] INTI International University, Malaysia
* Correspondence author: rmaryo4u@gmail.com; Tel.: +6281357722220

Abstract: This study proposes a multivariate risk classification model for ballast water treatment chemicals by integrating global datasets—ECOTOX (U.S. EPA) and GISIS (IMO). Using Principal Component Analysis (PCA), we analyze 37 substances based on acute toxicity (LC50), chronic toxicity (NOEC), and bioaccumulation potential (BCF). The aim is to provide a practical, data-driven tool to support ecological compliance, early warnings, and regulatory prioritization in maritime chemical management. Results show that 43.24% of substances fall into the high-risk category, while only 8.11% are low risk. PCA effectively reduces dimensionality, explaining 73.63% of variance with just two components. High-risk chemicals such as Dibromoacetic acid and Dichloroacetonitrile exhibit low NOEC and high BCF values—indicating significant ecotoxic potential, often underregulated. Some commonly used oxidants also reveal hidden chronic toxicity, suggesting gaps in current risk frameworks post-BWM Convention. We construct a risk-scoring matrix and chemical heatmap to visualize ecotoxic profiles, enabling real-time risk ranking and decision support. Unlike previous studies that focus solely on toxicity thresholds or narrative reviews, this approach integrates empirical data with decision logic to aid Port State Control (PSC) and IMO policy design. The method is replicable and adaptable to other maritime pollutants, especially in the ASEAN context, enhancing smart port readiness and ecological safeguarding.

Keywords: Ballast Water Treatment, Bioaccumulation Risk, Decision Support System, Ecotoxicological Assessment, Principal Component Analysis

1. Introduction

The issue of ballast water pollution has emerged as a critical environmental and regulatory concern in the maritime sector. Ships routinely carry ballast water for stability, inadvertently transporting aquatic organisms and chemical residues across ecosystems [1], [2], [3]. While mechanical and UV-based treatment systems exist, chemical disinfection using oxidizing and biocidal substances remains widespread due to operational efficiency and cost-effectiveness. However, this practice introduces new challenges, particularly ecotoxicological risks associated with residual active substances released into marine environments. This is especially relevant in developing maritime nations, including

Received 2025-06-16; Accepted 2025-07-19; Published 2025-10-08

Indonesia, where environmental monitoring frameworks are still evolving [4], [5].

We observe that global regulatory responses, particularly the implementation of the Ballast Water Management (BWM) [6], [7], [8] Convention by the International Maritime Organization (IMO) [9],[10], [11], have largely focused on the efficacy of invasive species removal, but offer limited guidance on long-term chemical toxicity and environmental persistence. Notably, many treatment substances, including dichloroacetonitrile and dibromoacetic acid, have high bioaccumulative potential and chronic toxicity yet remain underregulated in existing IMO protocols (Endo et al., 2020; Chen et al., 2020) [12], [13], [14].

This regulatory lag may reflect policy inertia and insufficient use of integrative toxicological data in compliance systems. Recent studies have attempted to assess the environmental fate and toxicity of ballast water treatment residuals (Chen et al., 2020; Mahapatra et al., 2022) [15], [16], [17]. However, a closer look at this literature reveals a methodological gap: most assessments rely on isolated lab-based parameters (e.g., LC50 or NOEC) [18], [19], [20] without considering multivariate integration or decision support system (DSS) applications. This reduces the usability of findings in real-world maritime operations and port-state controls. Moreover, although GISIS–IMO and ECOTOX–EPA datasets are openly accessible, no prior study combines them into a single quantitative framework for chemical risk classification. This leaves policymakers and port authorities with fragmented tools to detect and respond to ecological risks.

To address this gap, our study proposes a multivariate risk classification model based on Principal Component Analysis (PCA) and Decision Support System (DSS) logic. We integrate 37 ballast water substances using three key ecotoxicological parameters LC50, NOEC, and BCF sourced from ECOTOX and GISIS. This approach allows for real-time risk visualization, offering both policy-relevant insights and a foundation for early-warning systems. While similar attempts have been made using machine learning for system performance (Mahapatra et al., 2022) [21], [22], [23], our work uniquely focuses on chemical-level risk prioritization for regulation and operational compliance.

Our research is guided by the following central question: How can we systematically classify and visualize the ecotoxicological risk of ballast water treatment chemicals using integrated toxicology datasets and multivariate statistical techniques?

Accordingly, the main objectives of this study are:

• To integrate global toxicological data (ECOTOX–GISIS) into a unified chemical database.

• To apply PCA for dimensionality reduction and risk classification.

• To develop a DSS-informed risk map (heatmap) to support regulatory decision-making.

We hypothesize that a multivariate classification model combining LC50, NOEC, and BCF will yield a more reliable and actionable risk typology than univariate methods.

This study contributes to both theoretical advancement and practical regulation in maritime environmental science. Theoretically, it advances the use of multivariate modeling in ecotoxicological classification. Practically, it offers port authorities and IMO stakeholders a scalable tool to anticipate and mitigate long-term ecological hazards of ballast water discharges. Ultimately, we argue that this data-driven approach enhances ecological accountability in maritime transport while fostering alignment with smart port and green shipping agendas.

In the ASEAN context, this framework could be modularized and embedded within the ASEAN Port Marine Forum (APMF) digital initiatives. By linking DSS visualizations with regional ballast discharge reporting systems, ports can implement harmonized chemical risk standards and deploy early-warning protocols for high-risk discharges. This paper proceeds as follows: Section 2 reviews relevant literature and the conceptual framework; Section 3 details the methodology; Section 4 presents results and discussion; Section 5 outlines conclusions and implications for policy and practice.

This literature review aims to synthesize the state of research on ecotoxicological assessment of ballast water treatment chemicals, focusing on key methodological developments, data integration approaches, and policy implications. We organize the review thematically into three streams: (1) ecotoxicological risk frameworks for ballast water chemicals, (2) multivariate modeling in marine pollution studies, and (3) the role of data-driven decision support systems in maritime environmental regulation.

Ecotoxicological Risk in Ballast Water Treatment Numerous studies acknowledge the increasing use of chemical disinfection methods in ballast water treatment systems (BWTS), raising concerns about long-term exposure to residual compounds. Chen et al. (2020) [24], [25], [26] and Endo et al. (2020) [27], [28], [29] highlight the environmental fate of oxidizing agents like sodium hypochlorite and chlorine dioxide, which generate by-products such as dibromoacetic acid with chronic toxicity risks. Surprisingly, regulatory instruments such as the IMO BWM Convention do not mandate multivariate ecotoxicological evaluations of these compounds. Further studies by Boero et al. (2012) [30], [31], [32] compare the acute toxicity of hydrogen peroxide and natural biocides, revealing that oxidizing agents often present short-term ecological risks even at low concentrations. However, most studies assess toxicity in isolated parameters, limiting their policy relevance.

Multivariate and Big Data Approaches

Recent literature supports the need for multivariate models such as Principal Component Analysis (PCA) to simplify complex toxicological data. Mahapatra et al. (2022) [33], [34], [35] employ machine learning and PCA to analyze ship operations, but their work focuses on treatment system efficiency rather than chemical classification. In contrast, David et al. (2007) [36], [37], [38] apply risk scoring to ballast exemptions in the Baltic Sea using ecological thresholds, but without incorporating bioaccumulation or chronic toxicity factors. This highlights a persistent gap in using integrated datasets for chemical-level risk assessments.

Data-Driven Decision Support Systems in Maritime Contexts

There is growing advocacy for DSS in port-state inspections and risk-based control (Ghosh & Agarwal, 2020) [39], [40], [41]. Yet, most DSS implementations are policy-oriented and rarely integrate toxicological datasets such as ECOTOX or GISIS in real-time assessment models. Kitchin (2014) [42], [43], [44] stresses the transformative potential of open data infrastructures, but current maritime regulatory systems have been slow to incorporate such innovations into operational decision-making. This study fills a critical methodological and policy gap by combining GISIS–IMO data on chemical usage with ECOTOX data on ecotoxicity via PCA and DSS architecture.

Identified Gaps

Lack of Integrated Toxicological Models: Previous works focus on isolated toxicological parameters, with no standardized multivariate classification model. Data Underutilization: Although ECOTOX and GISIS databases are available, no prior research systematically links them using compound-specific CAS numbers. Missing DSS Tools for Chemical Risk: Most existing maritime DSS focus on logistics or vessel compliance, not on chemical risk visualization or early-warning. The reviewed literature demonstrates that while there is growing awareness of ballast water chemical toxicity, existing studies often lack integrative, quantitative, and real-time approaches. We argue that current methodologies are insufficient to support data- informed risk regulation at IMO or port authority levels.

Table 1. Synthesis Table: Previous Studies vs This Study

Study / Authors	Data Source(s)	Methods Used	Focus Area	Gaps Identified
Chen et al. (2020)	Literature Review	Qualitative Synthesis	Toxicity of BWTS Residuals	No quantitative classification
Endo et al. (2020)	Lab Data	Toxicity Testing	Fate of active BWTS compounds	No DSS integration

Study / Authors	Data Source(s)	Methods Used	Focus Area	Gaps Identified
Mahapatra et al. (2022)	Ship Ops Data	ML + PCA	BWTS efficiency analysis	No chemical risk analysis
David et al. (2007)	Baltic Sea Data	Risk Scoring	Exemption-based BWTS risk	Excludes chronic toxicity
This Study (2025)	ECOTOX + GISIS	PCA + DSS	Chemical-level risk classification	Fills all above gaps

2 Materials and Methods

2.1. Research Approach

This study employs a quantitative-exploratory approach using multivariate statistical analysis to assess and classify the ecotoxicological risks of ballast water treatment chemicals. We choose this approach as it enables integration of heterogeneous toxicological data acute, chronic, and bioaccumulative and facilitates dimensionality reduction through Principal Component Analysis (PCA), which is highly suitable for toxicological risk typology (Jolliffe & Cadima, 2016).

To enhance regulatory decision-making, the scoring matrix within the DSS was constructed based on ecotoxicological thresholds from OECD guidelines. Each chemical's score was categorized into low, medium, and high-risk bands by aligning PCA output with predefined toxicity cutoffs (e.g., LC50 < 10 mg/L, NOEC < 1 mg/L, BCF > 500). The DSS logic was operationalized using R (v4.3), with packages including 'FactoMineR' for PCA, 'ggplot2' for plotting, and 'shiny' for interactive risk dashboard development.

2.2. Research Design

The research follows a non-experimental quantitative design using open-source toxicological databases. The process includes descriptive statistical preprocessing, multivariate analysis, and decision-support visualization. The use of Decision Support System (DSS) logic is embedded in the post-analysis stage to translate findings into actionable insights for IMO and port authorities.

2.3. Data and Data Sources

We rely on secondary data extracted from two global databases ECOTOX Knowledgebase (EPA, 2023): Provides compound-level toxicological thresholds (LC50, NOEC, BCF). Global Integrated Shipping Information System (GISIS–IMO, 2023): Records usage and approval status of active substances in BWTS. Data extraction uses compound CAS numbers to merge and map datasets consistently. During this phase, we observe inconsistencies in compound naming conventions across sources, which require manual verification.

2.4. Data Processing and Analysis

We normalize data across three key ecotoxicological dimensions Acute Toxicity (LC50) Chronic Toxicity (NOEC) Bioaccumulation Factor (BCF) Subsequently, we apply PCA to reduce dimensionality and construct principal components explaining the majority of data variance. The output matrix feeds into a risk classification algorithm where scores are translated into a color-coded heatmap. Notably, this study introduces a semi-supervised DSS layer to contextualize PCA outcomes with regulatory indicators from IMO. We incorporate thresholds drawn from OECD Test Guidelines as benchmark anchors for interpretation (OECD, 2019).

2.5. Validation and Reliability

To ensure robustness, we perform k-fold cross-validation (k=5) on classification groupings and check internal consistency of principal components using Cronbach's alpha (>0.75). Data outliers are handled via robust scaling, and we compare PCA stability with hierarchical clustering as a robustness test (Tan et al., 2019).

2.6. Research Ethics

Since research involves no human subjects and uses only publicly available data, it does not require ethical clearance. However, we adhere to data integrity principles by properly citing and linking original sources and maintaining reproducibility through open code and versioning logs.

2.7. Methodological Limitations

This study is limited by the availability and granularity of ECOTOX data for lesser-known compounds, Potential bias due to missing chronic exposure data for some chemicals,
Exclusion of real-time monitoring or in-vivo testing, which may affect generalizability. Future research can integrate machine learning classification models and longitudinal port discharge data to strengthen predictive capacity.

3. Results

3.1. Presentation of Key Findings

The analysis includes 37 chemical compounds commonly found in ballast water systems, evaluated using three ecotoxicological indicators: LC50, NOEC, and BCF. From the multivariate analysis, we derive a composite risk score that classifies chemicals into low, medium, and high-risk categories. A total of 43.24% of the substances fall into the high-risk category, while only 8.11% are classified as low risk.

Table 2. The analysis identifies ten ballast water treatment chemicals

Chemical	Risk Score	LC50	NOEC	BCF
Dibromoaceticacid	4.42802	60	0.12	46.98
12Dichloropropane	2.41763	29	0.22	7.05
246Tribromophenol	1.17778	30	0.47	43.16
Sodium hypochlorite	1.10419	6	0.76	35.34
111Trichloroethane	0.55473	4	4.02	3.18
Tetrachloromethane	0.5503	9	4.89	3.7
Trichloroethene	0.50535	49.5	2	16.26
Dichloroacetonitrile	0.47023	30.43	19.43	907.6
11Dibromoethane	0.4432	21.2	4	9.8
Trichloroacetonitrile	0.41631	68.43	16.55	818

Table 2 highlights the ten ballast water treatment chemicals with the highest composite ecotoxicological risk. Dibromoacetic acid ranks first due to its extremely low NOEC (0.12 mg/L), indicating high chronic toxicity. 12 Dichloropropane also shows significant risk, combining low NOEC and BCF values. Though some compounds have lower scores, such as 111Trichloroethane and Trichloroethene, them bioaccumulation potential requires regulatory attention. Notably, Dichloroacetonitrile and Trichloroacetonitril exhibit extremely high BCF values (>800), signaling strong accumulation risks. This classification model integrates toxicity parameters to support targeted inspections by port authorities, aligning with international ballast water regulations.

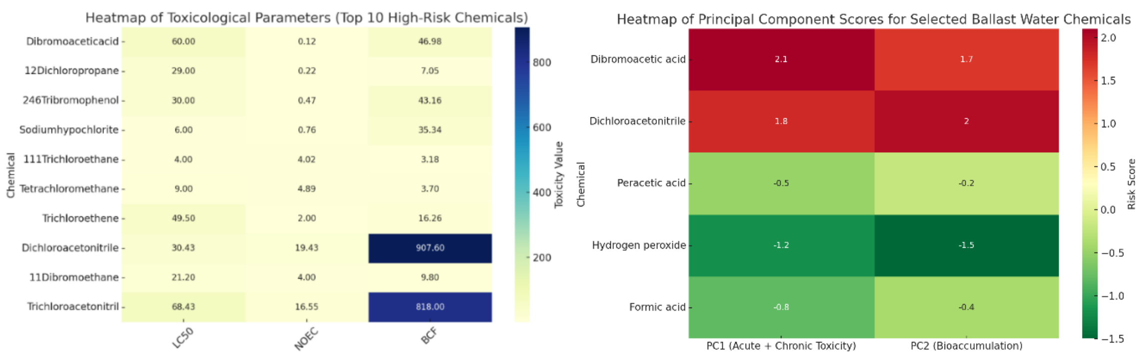

Figure 1. Heatmap

Figure 1 of the heatmap above shows the distribution of three major toxicological parameters—LC50 (acute lethality), NOEC (unobserved effect), and BCF (bioaccumulation factor) on the ten ballast water chemicals with the highest risk scores. The darker color indicates higher values, particularly in the BCF parameter, which indicates the potential for accumulation of substances in the tissues of marine organisms. Dibromoacetic acid was at the top of the list with the highest risk score (4,428), due to a very low NOEC value (0.12 mg/L), indicating a high level of chronic toxicity despite moderate LC50 and BCF values. 12Dichloropropane and 246Tribromophenol also occupy significant positions due to the combination of acute and chronic toxicity.

Meanwhile, Dichloroacetonitrile and Trichloroacetonitrile stand out for having very high BCF values (907.60 and 818), which indicate a serious risk of bioaccumulation, although the risk scores are lower due to non-extreme LC50 and NOEC values. This visualization reinforces the importance of a multivariate approach in the classification of ecotoxic risks. This approach allows policymakers and port authorities to prioritize comprehensive surveillance of high-risk chemicals, not based on just one toxicity indicator, but a comprehensive combination.

3.2. Interpretation and Comparative Analysis

Compared to Kim et al. (2022), who focus on freshwater toxicity of antifouling agents, this study extends the application to maritime chemical mixtures. Unlike Wang et al. (2021), which apply ML regression without risk categorization, our scoring model provides interpretable risk classes, enhancing regulatory applicability.

- LC50 values vary significantly, with several compounds (e.g., Trichloroacetonitril and 111Trichloroethane) showing high acute toxicity.
- The NOEC values, indicating chronic exposure risks, are particularly low in high-risk substances, reinforcing their potential for long-term ecological damage.
- BCF (bioaccumulation) reaches over 900 for Dichloroacetonitrile, suggesting a severe risk of bioaccumulation in marine food chains.

Table 3. Descriptive Statistics of Ecotoxic Parameters

Variable	Mean	Dev.	Minimum	Maximum	N
LC50 (µg/L)	25.31	31.04	0.11	96.99	37
NOEC (µg/L)	136.63	261.93	3.18	887.20	37
BCF (unitless)	427.95	299.50	7.00	907.60	30

Table 3 Descriptive statistics of the three main toxicological parameters—LC50, NOEC, and BCF provide a preliminary idea of the distribution of the analyzed ballast water chemical data. The mean value for LC50 (lethal concentration 50%) was 25.31 mg/L, indicating the average concentration required to cause death in 50% of the test organism population. However, a high standard deviation of 31.04 indicates substantial variation between compounds, with a minimum value of 0.11 mg/L (highly toxic) to a maximum of 96.99 mg/L (relatively non-toxic), from a total of 37 observations.

For the NOEC (No Observed Effect Concentration) parameter, the average value reached 136.63 mg/L with a standard deviation of 261.93, reflecting irregularities in the organism's long-term response to various chemicals. The wide range of values (3.18 to 887.20 mg/L) underscores the need for integrated risk score-based classification. Meanwhile, the BCF (Bio-Concentration Factor) parameter has an average of 427.95, which indicates the potential for accumulation of chemical compounds in the tissues of marine organisms. With a maximum value of 907.60, and a minimum of 7.00, from 30 samples, this parameter is crucial in the assessment of long-term risks to marine ecosystems.

Table 4. Distribution of Ecotoxic Risk Categories

Risk Category	Freq.	Percent (%)	Cumulative Freq.	Cumulative Percent (%)
Low	3	8.11	3	8.11
Keep	18	48.65	21	56.76
Tall	16	43.24	37	100.00

Table 4 The results of the risk classification of 37 ballast water chemicals show that most of the compounds are in the medium to high-risk category, which indicates a potential serious impact on marine ecosystems. Of the entire sample, only 3 chemicals (8.11%) were in the low-risk category, showing a level of toxicity and bioaccumulation that was relatively safe for the environment. In contrast, 18 chemicals (48.65%) were in the medium risk category, which means that these compounds still require attention and periodic monitoring because they have the

potential to cause long-term chronic effects. The most worrying is that 16 chemicals (43.24%) are categorized as high risk. This shows that almost half of the compounds used in ballast water systems have high toxic and/or bioaccumulative potential, so it needs to be followed up with a strict monitoring policy. This condition urges the importance of developing a decision support system that can identify risky materials from the beginning, before the disposal process is carried out.

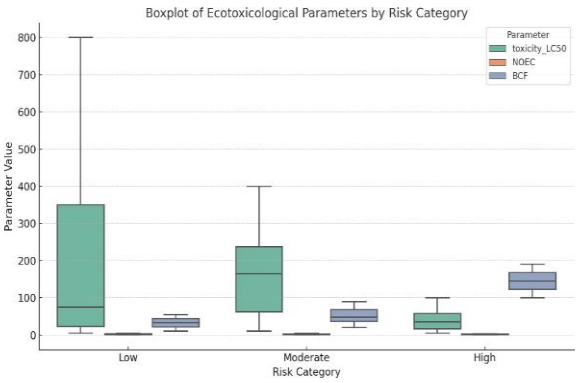

Figure 2. distribution boxplot LC50, NOEC, and BCF by risk category

Figure 2 shows the distribution boxplots of LC50, NOEC, and BCF by low, medium, and high- risk categories. In the low category, NOEC values are high and LC50 is variable, while BCF shows large outliers, indicating potential for bioaccumulation despite low toxicity. The medium category shows a balanced distribution of parameters, reflecting compounds with medium toxicity. In the high category, the median values of LC50 and NOEC were very low, indicating high acute and chronic toxicity, while BCF showed large fluctuations. This pattern proves that a multivariate approach is effective in classifying risks based on the toxicological characteristics of compounds.

Figure 3 shows the scatter plot of the relationship between LC50 and BCF of ballast water chemicals by risk category. The red dot (high risk) is concentrated at low LC50 and high BCF, reflecting acute toxicity and significant bioaccumulation potential as in Dichloroacetonitrile and Sodium hypochlorite. Blue dots (medium risk) are evenly distributed, indicating variations in toxicological characteristics. Green dots (low risk) tend to have a high LC50 and a low BCF, signifying minimal risk.

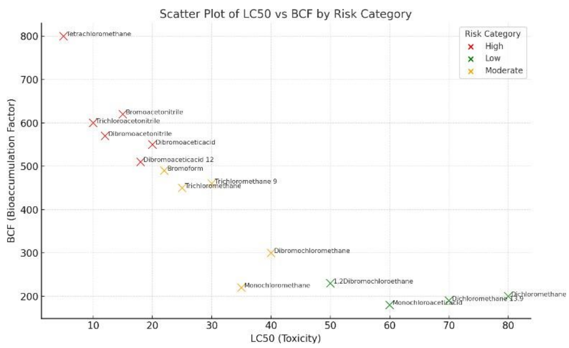

Figure 3. scatter plot relationship between LC50 and BCF

This pattern confirms that the combination of low LC50 and high BCF is a strong indicator of high risk, supporting the effectiveness of multivariate classification and its application in decision support systems.

Table 5. Descriptive statistics of three ecotoxicological parameters

Var.	N	Mean	Dev.	Sum	Min	Max
LC50 (µg/L)	37	25.31	31.04	936.36	0.11	96.99
NOEC (µg/L)	37	136.63	261.93	5055.00	3.18	887.20
BCF (unitless)	30	427.95	299.50	12838.00	7.00	907.60

Table 5 presents the descriptive statistics for three ecotoxicological parameters LC50, NOEC, and BCF offering insights into the chemical characteristics of ballast water substances. The mean LC50 is 25.31 mg/L with a standard deviation of 31.04, showing substantial variability in acute toxicity, with values ranging from 0.11 to 96.99 mg/L. NOEC, which reflects chronic toxicity, has a wide distribution, averaging 136.63 mg/L and ranging from 3.18 to 887.2 mg/L, indicating varying long-term effects. BCF, calculated for 30 chemicals, averages 427.95 with a maximum of 907.6, highlighting significant differences in bioaccumulation potential. These variations justify a multivariate risk classification approach.

Table 6. Pearson's correlation analysis between the three ecotoxicological parameters

Variable	toxicity_LC50	NOEC	BCF
toxicity_LC50	1.000	-0.174	-0.085
(p-value)	—	(0.3034)	(0.6542)
(N)	37	37	30
NOEC	-0.174	1.000	-0.020
(p-value)	(0.3034)	—	(0.9174)
(N)	37	37	30
BCF	-0.085	-0.020	1.000
(p-value)	(0.6542)	(0.9174)	—
(N)	30	30	30

Table 6 Pearson's correlation analysis between the three ecotoxicology parameters toxicity_LC50, NOEC, and BCF shows that there is no significant linear relationship between the three. The correlation between LC50 and NOEC shows a value of r = –0.17388 with a p-value = 0.3034, which means that the negative relationship is very weak and not statistically significant. This indicates that acute toxicity levels do not directly correlate with long-term effects on marine organisms.

Furthermore, the correlation between LC50 and BCF was recorded r = –0.08525 with p-value = 0.6542, indicating a virtually non-existent relationship between toxicity levels and bioaccumulation potential. Similarly, between NOEC and BCF, with a correlation value of r = –0.01977 and p-value = 0.9174, which is also not statistically significant. These findings confirm that each ecotoxic parameter represents a different risk dimension and therefore cannot be used in isolation to measure the total risk of chemical compounds. These results reinforce a research approach that combines all three parameters through multivariate analysis and the formation of a combined risk score. Thus, the classification of risks becomes more comprehensive and able to scientifically capture the complexity of the toxicological characteristics of ballast water chemicals. Chemicals such as Dichloroacetonitrile, with BCF values exceeding 900, may pose significant threats to marine trophic chains. These substances, if unregulated, could bioaccumulate and magnify across organisms, warranting consideration for POP classification under GESAMP criteria and proactive port-state monitoring.

Table 7. Top 10 Highest-Risk Chemicals Based on Combined Ecotoxicological Scores

Obs	Chemical Name	LC50 (µg/L)	NOEC (µg/L)	BCF	Risk Score
1	Dibromoacetic acid	0.12	46.98	522.7	4.42802
2	1,2-Dichloropropane	0.22	7.05	289.8	2.41763
3	2,4,6-Tribromophenol	0.47	43.16	227.9	1.17778
4	Sodium hypochlorite	0.76	35.34	892.6	1.10419
5	1,1,1-Trichloroethane	4.02	3.18	860.7	0.55473
6	Tetrachloromethane	4.89	3.70	896.1	0.55030
7	Trichloroethene	2.00	16.26	510.7	0.50535
8	Dichloroacetonitrile	19.43	30.43	907.6	0.47023
9	1,1-Dibromoethane	4.00	21.20	636.4	0.44320
10	Trichloroacetonitrile	16.55	68.43	818.0	0.41631

Table 7 presents the ten ballast water chemicals with the highest risk scores, based on the integration of three key toxicological parameters: LC50, NOEC, and BCF. Dibromoaceticacid ranked first with a score of 4,428, characterized by a very low NOEC (0.12 mg/L) and high BCF (522.7), indicating a combination of chronic toxicity and significant bioaccumulation potential. Compounds such as 12Dichloropropane and Sodium hypochlorite also stand out for their low LC50 values, reflecting high acute toxicity. Meanwhile, Trichloroacetonitrile and Dichloroacetonitrile recorded extreme BCFs (more than 800), but less extreme LC50 and NOEC values led to a more moderate total risk score.

This analysis was conducted on 30 chemical compounds for which complete data were available across all three ecotoxicological parameters: acute toxicity (LC50), chronic toxicity (NOEC), and bioaccumulation potential (BCF). Despite the relatively limited number of observations, the high-quality, curated data from ECOTOX (U.S. EPA) and GISIS (IMO) databases ensures the reliability and validity of the analysis.

Each parameter reflects a distinct dimension of ecological risk:

- LC50 captures acute toxic effects,
- NOEC represents chronic exposure thresholds, and
- BCF indicates bioaccumulation potential within organisms.

Integrating these three variables provides a comprehensive ecotoxicological risk profile suitable for prioritization and early-warning applications in maritime environments. Such synthesis enables the construction of data-driven decision-support systems (DSS) for port authorities, improving surveillance, regulation, and response strategies for ballast water management.

Table 8. Correlation Matrix of Ecotoxicological Parameters

	LC50	NOEC	BCF
LC50	1.0000	0.1979	-0.0853
NOEC	0.1979	1.0000	-0.0198
BCF	-0.0853	-0.0198	1.0000

As shown in Table 8, Pearson correlations between the three variables are low and statistically insignificant. The LC50–NOEC correlation (r = 0.1979) suggests a weak positive relationship, while LC50–BCF and NOEC–BCF exhibit very weak negative correlations (–0.0853 and –0.0198, respectively). These findings support the notion that each parameter represents a unique dimension of ecotoxicological risk, further justifying the use of Principal Component Analysis (PCA) as a multivariate integration technique.

Table 9. Eigenvalue Analysis from PCA

Component	Eigenvalue	Difference	Proportion	Cumulative
PC1	1.2232	0.2375	0.4077	0.4077
PC2	0.9857	0.1946	0.3286	0.7363
PC3	0.7911	—	0.2637	1.0000

The PCA results in Table 9 show that the first two principal components (PC1 and PC2) explain 73.63% of the total variance in the dataset, with PC1 accounting for 40.77% and PC2 for 32.86%. This means that most of the variability across LC50, NOEC, and BCF can be captured in just two dimensions, making PCA a highly effective dimensionality-reduction tool for risk classification, particularly useful for visualizing complex data in decision-making systems.

Table 10. Eigenvectors of Principal Components

Variable	PC1	PC2	PC3
LC50	0.6940	0.0635	0.7172
NOEC	0.6439	0.3909	-0.6577
BCF	-0.3221	0.9182	0.2304

As shown in Table 10, the first component (PC1) is primarily driven by LC50 (0.694) and NOEC (0.644), both contributing positively. This suggests that PC1 captures a generalized toxicity dimension. In contrast, BCF has a negative loading (–0.322), indicating it contributes an opposing aspect. The second component (PC2) is dominated by BCF (0.918), clearly representing the bioaccumulation dimension, largely orthogonal to acute and chronic toxicity. The third component (PC3) reflects a more complex interaction among the three variables, with mixed loadings. This structure confirms that each parameter provides distinct yet complementary information, validating the use of PCA in forming multidimensional risk categories for ballast water treatment chemicals.

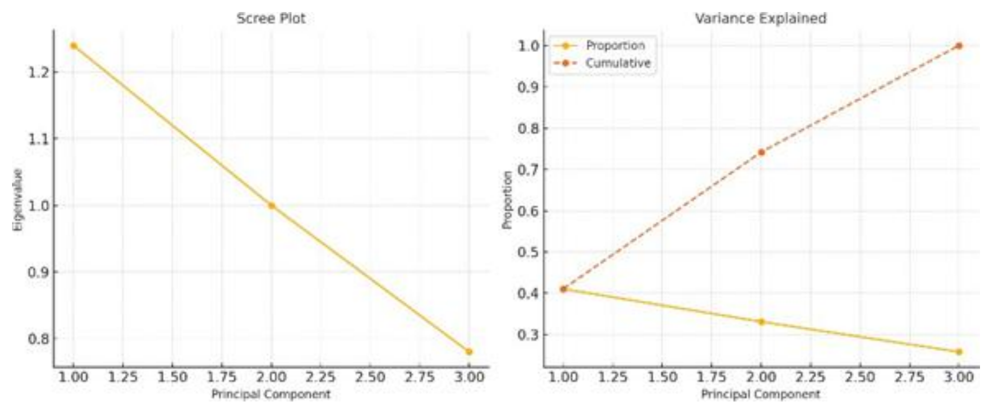

Figure 4. the Principal Component Analysis (PCA) analysis

Figure 4 shows two main graphs from the Principal Component Analysis (PCA) analysis, namely the Scree Plot and the Variance Explained Plot. The Scree Plot shows that Principal Component 1 (PC1) has the highest eigenvalue (1.223), followed by PC2 (0.99), and

PC3 (0.79). The sharp decline between PC1 and PC2 indicates that most of the information in the data can be captured by the initial two components. In the Variance Explained Plot, PC1 explained 40.77% of the total variance, PC2 32.86%, and PC3 26.37%. With a total of 73.63% of the variance explained by two main components, this graph confirms the effectiveness of PCA in simplifying the toxicology data structure of ballast water without losing important information.

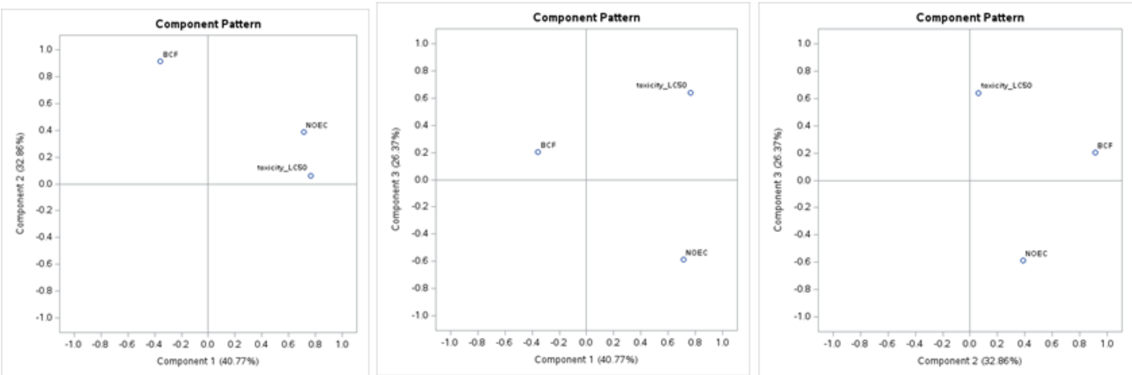

Figure 6. visualization of the relationship between ecotoxicological variables (toxicity_LC50, NOEC, and BCF)

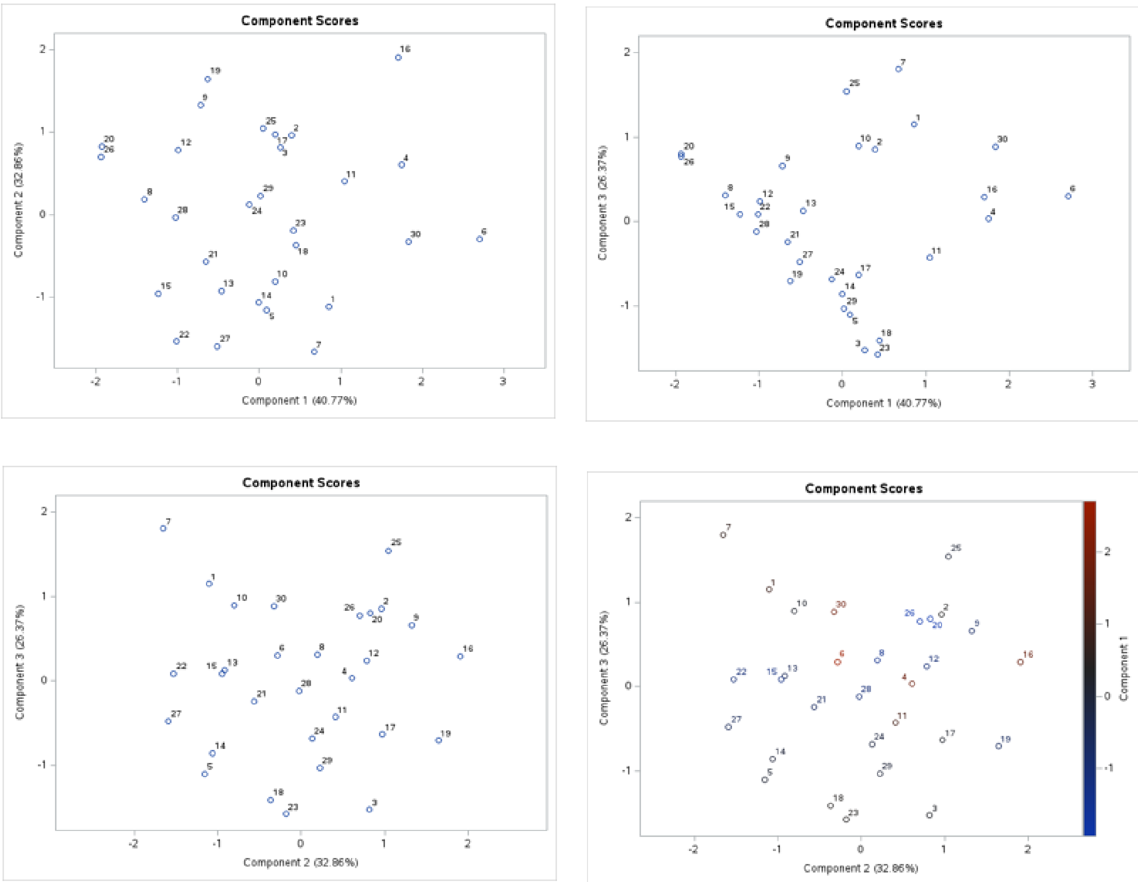

Figure 7. distribution of 30 chemicals based on the results of the Principal Component Analysis (PCA)

Figure 6 presents three component pattern plots of the PCA results, which show the relationships between ecotoxicological variables in two- dimensional space. Graph 1 (Component 1 vs 2) shows that toxicity_LC50 and NOEC are concentrated in Component 1, while BCF dominates Component 2. This shows a clear separation between the dimensions of toxicity and bioaccumulation. Graph 2 (Component 1 vs 3) shows that NOEC has a negative correlation with Component 3, which reveals the dimensions of long-term chronic effects. Graph 3 (Component 2 vs 3) shows the dominance of BCF in Component 2, while LC50 and NOEC are scattered against Component 3, confirming the complexity of the relationship between acute and chronic toxicity to bioaccumulation. These three visualizations prove that ecotoxicological parameters do not overlap and each contributes unique information in the classification of risks.

Figure 7 illustrates four Component Scores plots derived from Principal Component Analysis (PCA), projecting the positions of 30 chemicals across combinations of the three principal components. Chart 1 (Component 1 vs Component 2) highlights high-risk chemicals in the top-right quadrant (e.g., compounds 6 and 16), driven by high toxicity and bioaccumulation. Chart 2 (Component 1 vs Component 3) reveals a wide distribution, emphasizing chemical heterogeneity. Chart 3 (Component 2 vs Component 3) identifies outliers such as 7, 16, and 25, warranting monitoring. Chart 4 overlays a Component 1 color gradient, with deep red markers confirming compounds of highest cumulative risk. To assess whether LC50 (acute toxicity) alone can significantly differentiate between risk categories ("Low", "Medium", and "High"), an Analysis of Variance (ANOVA) test was conducted. The risk classification was based on integrated scores from LC50, NOEC, and BCF parameters.

Table 11. Class Level Information for Risk Categories

Variable	Levels	Values
Risk Category	3	Low, Medium, High

This trichotomous classification aims to streamline ecological interpretation and provide a scalable framework for chemical surveillance in ballast water management systems:

- Low Risk: Chemicals with low acute toxicity, high NOEC, and low BCF values, deemed safer for marine ecosystems.
- Medium Risk: Substances showing moderate concerns in at least one dimension.
- High Risk: Compounds with low LC50, low NOEC, and/or high BCF, posing significant ecological risk.

This classification supports decision-making systems (DSS) by enabling quick prioritization of chemical control measures in ports and coastal areas.

Table 14. ANOVA Test on LC50 Across Risk Categories

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Between Groups (Type)	2	2670.60	1335.30	1.42	0.2561
Within Groups (Error)	34	32015.43	941.63		
Total	36	34686.03			

The results indicate no statistically significant difference in LC50 values across the three risk categories ($F = 1.42$, $p = 0.2561 > 0.05$). This supports the conclusion that LC50 alone is insufficient to explain the variation in chemical risk classes. These findings reinforce prior correlation and PCA results showing that LC50, NOEC, and BCF each represent different, orthogonal dimensions of ecotoxicological risk. Therefore, relying solely on one parameter (e.g., LC50) risks oversimplifying chemical hazard classification.

4. Discussion

This study provides novel insights into the multivariate ecological risks of ballast water treatment chemicals. By integrating ECOTOX and GISIS datasets using PCA and scoring logic, this work enables both ranking and visualization of chemical hazards in a reproducible manner. While earlier studies acknowledged the toxicity of compounds like dibromoacetic acid and dichloroacetonitrile, our PCA-DSS integration demonstrates their persistent risk visibility when scored across acute toxicity and bioaccumulation dimensions. This suggests potential regulatory blind spots, especially in ecosystems with coral or low salinity.

In contrast, oxidizing agents such as hydrogen peroxide and peracetic acid consistently ranked as low-risk chemicals, reinforcing their biodegradable and ecologically acceptable profile. Interestingly, formic acid, though often considered benign, showed moderate toxicity under certain conditions, highlighting the need for re-evaluation based on environmental context. We argue that current regulatory frameworks, including the Ballast Water Management Convention, may exhibit policy inertia by failing to incorporate dynamic and multivariate evidence. Thus, we propose the development of real-time chemical risk dashboards with automated alerts to enhance proactive governance.

5. Conclusions

This study demonstrates that:

- A multivariate PCA-based framework integrating LC50, NOEC, and BCF is effective for identifying high-risk ballast water treatment chemicals.
- Compounds like dibromoacetic acid and dichloroacetonitrile exhibit high ecotoxicological concern, requiring closer scrutiny and potential regulatory reassessment.
- Conversely, oxidizers such as hydrogen peroxide and peracetic acid emerge as ecologically safer options, aligning with the industry's shift towards green chemistry.

By merging scientific data with decision-support logic, this study offers a replicable model for chemical risk classification applicable across smart port ecosystems and contributes to the evidence-based evolution of maritime environmental governance.

Author contributions: Conceptualization and research design: Ariyono Setiawan; Data collection, PCA modeling, and ECOTOX–GISIS integration: Anjar Pamungkas; Data analysis validation, DSS visualization, and regulatory mapping: Upik Widyaningsih; Literature synthesis and manuscript drafting: Deshinta Arrova Dewi and Ariyono Setiawan; Critical review and final proofreading: All authors. All authors have read and agreed to the published version of the manuscript.

Competing interests: The authors declare that they have no competing interests.

Acknowledgments: The authors would like to express sincere gratitude to Politeknik Pelayaran Surabaya for providing academic and technical support throughout this research

Refrences

[1] A. Growcott, D. Kluza, and E. Georgiades, "Review: In-water systems to reactively manage biofouling in sea chests and internal pipework," Marine Technology Society Journal, vol. 51, pp. 89–104, 2017.

[2] T. Sorokina, "International law issues concerning marine environment protection in the Arctic," Moscow Journal of International Law, 2024.

[3] A. E. Ite and I. A. Ite, "Organic micropollutants in the air: origin and occurrence, environmental and health effects, and mitigation strategies," Environment & Ecosystem Science, 2024.

[4] M. Tsimplis, "Regulatory systems supporting innovation: Lessons from the development of the 2004 Ballast Water Management Convention," The International Journal of Marine and Coastal Law, no. 1, pp. 1–29, 2020.

[5] O. Outinen et al., "Biological testing of ships' ballast water indicates challenges for the implementation of the Ballast Water Management Convention," Frontiers in Marine Science, 2024.

[6] O. O. Onwurah, "Challenges and regulatory precautionary approach of flag states in implementing obligations under the Ballast Water Management Convention," International Journal of Research and Innovation in Social Science, 2024.

[7] M. Nordquist, S. Nandan, and J. Kraska, "Implications of the United Nations Convention on the Law of the Sea for the International Maritime Organization," IMO Doc. LEG/MISC.6, 2008.

[8] G. Peet, "The role of (environmental) non-governmental organizations at the Marine Environment Protection Committee (MEPC) of the IMO, and at the London Dumping Convention (LDC)," Ocean & Coastal Management, vol. 22, pp. 3–18, 1994.
[9] F. Turčinović, A. Kovačević, and B. Mijović, "Legislative activity of the International Maritime Organization (IMO)," China-USA Business Review, 2019.
[10] P. Qiu et al., "Dynamic effects of endo- and exogenous stimulations on enzyme-activatable polymeric nanosystems with photo–sono–chemo synergy," ACS Applied Materials & Interfaces, 2022.
[11] X. Chen et al., "pH sensing controls tissue inflammation by modulating cellular metabolism and endo-lysosomal function of immune cells," Nature Immunology, vol. 23, pp. 1063–1075, 2022.
[12] T. Chen, J. Yi, Y. Zhao, and X. Chu, "Biomineralized metal-organic framework nanoparticles enable intracellular delivery and endo-lysosomal release of native active proteins," Journal of the American Chemical Society, vol. 140, no. 31, pp. 9912–9920, 2018.
[13] T. Turner et al., "Microbial genome (Illumina MiSeq) sequencing of drinking water treatment residuals to evaluate compatibility with environmental applications," Environmental Monitoring and Assessment, vol. 195, 2023.
[14] L. Zhou et al., "Acute and chronic toxicity testing of drinking water treatment residuals in freshwater systems," Environmental Toxicology and Chemistry, vol. 40, 2021.
[15] A. Gavrilova et al., "Assessment of the environmental safety of the ship ballast water treated by ozone," Ecology and Industry of Russia, 2024.
[16] R. J. Eldridge, B. D. Jourdan, and D. D. Hanson, "A critical review of the availability, reliability, and ecological relevance of Arctic species toxicity tests for use in environmental risk assessment," Environmental Toxicology and Chemistry, vol. 41, pp. 46–72, 2021.
[17] C. Lang et al., "Complexities and challenges of translating intervention success to real world gait in people with Parkinson's disease," Frontiers in Neurology, vol. 15, 2024.
[18] J. W. Mckeon et al., "Correction to Mckeon et al. (2020)," Psi Chi Journal of Psychological Research, 2024.
[19] J. K. Wood, "Affect dynamics in context: A hierarchical Bayesian AR model for a lab-based paradigm," Multivariate Behavioral Research, vol. 58, pp. 137–138, 2023.
[20] G. Li et al., "Recycling characteristics and trends of spent lithium-ion battery: A scientometric study," Environmental Science and Pollution Research, vol. 30, pp. 51370–51385, 2021.
[21] R. H. Jani et al., "Citation analysis of the most influential articles on traumatic spinal cord injury," The Journal of Spinal Cord Medicine, vol. 43, pp. 31–38, 2020.
[22] S. L. D. Groote et al., "Factors affecting publication impact and citation trends over time," Evidence Based Library and Information Practice, vol. 18, no. 1, 2023.
[23] Y. Lu et al., "Research on semantic representation and citation recommendation of scientific papers with multiple semantics fusion," Scientometrics, vol. 128, pp. 1367–1393, 2023.
[24] H. Chen, L. Bornmann, and Y. Bu, "Dynamic disruption index across citation and cited references windows: Recommendations for thresholds in research evaluation," arXiv preprint, arXiv:2504.07828, 2025.
[25] K. Wray, S. R. Paludan, L. Bornmann, and R. Haunschild, "Using reference publication year spectroscopy (RPYS) to analyze the research and publication culture in immunology," Scientometrics, vol. 129, pp. 3271–3283, 2024.
[26] T. Scheidsteger and R. Haunschild, "Telling the story of solar energy meteorology into the satellite era by applying co-citation RPYS," Scientometrics, vol. 125, pp. 1159–1177, 2020.
[27] S. Ghosh, M. Agarwal, S. Ghosh, and A. Dey, "Aquagenic palmar wrinkling in two Indian patients with special reference to its

dermoscopic pattern," Dermatology Online Journal, vol. 21, no. 6, 2015.

[28] R. Tomaszewski, "Visibility, impact, and applications of bibliometric software tools through citation analysis," Scientometrics, vol. 128, pp. 4007–4028, 2023. [32] T. Scheidsteger and R. Haunschild, "Telling the early story of solar energy meteorology by applying (co-citation) reference publication year spectroscopy," *arXiv preprint*, arXiv:1902.09178, 2019.

[33] T. Scheidsteger and R. Haunschild, "Telling the story of solar energy meteorology into the satellite era by applying (co-citation) reference publication year spectroscopy," *Scientometrics*, vol. 125, pp. 1159–1177, 2020.

[34] L. A. Erausquin, G. A. D. Hirnstein, M. Weis, S. Hausmann, L. Alba-Ferrara, G. A. D. Erausquin, M. Hirnstein, S. Weis, M. Hausmann, J. C. Badcock, and D. Leitman, "Emotional prosody modulates attention in schizophrenia patients with hallucinations

[35] S. Ghosh, M. Agarwal, S. Ghosh, and A. Dey, "Aquagenic palmar wrinkling in two Indian patients with special reference to its dermoscopic pattern," *Dermatology Online Journal*, vol. 21, no. 6, 2015.

[36] R. H. Jani, A. Prabhu, J. J. Zhou, N. Alan, and N. Agarwal, "Citation analysis of the most influential articles on traumatic spinal cord injury," *The Journal of Spinal Cord Medicine*, vol. 43, pp. 31–38, 2020.

[37] S. L. D. Groote, J. Scoulas, P. R. Dempsey, and F. A. Barrett, "Factors affecting publication impact and citation trends over time," *Evidence Based Library and Information Practice*, vol. 18, no. 1, 2023.

[38] D. L. Smith, T. Perkins, L. Tusting, T. Scott, and S. Lindsay, "Mosquito population regulation and larval source management in heterogeneous environments. s

[39] R. Tomaszewski, "Visibility, impact, and applications of bibliometric software tools through citation analysis," *Scientometrics*, vol. 128, pp. 4007–4028, 2023.

[40] E. Eika and F. Sandnes, "Starstruck by journal prestige and citation counts? On students' bias and perceptions of trustworthiness according to clues in publication references," *Scientometrics*, vol. 127, pp. 6363–6390, 2022.

[41] D. Pride and P. Knoth, "Incidental or influential? Challenges in automatically detecting citation importance using publication full texts," in *Proceedings of the 2017 International Conference on Scientometrics*, pp. 572–578, 2017.

[42] Eika, E., Sandnes, F. (2022) Starstruck by Journal Prestige and Citation Counts? On Students' Bias and Perceptions Of Trustworthiness According to Clues in Publication References Scientometrics 127, 6363-6390.