Optimised Ship Fuel Prediction for meeting Sulphur Abatement Standards in Shipping Industry

Yuting Hu
School of Computing
University of Portsmouth
Portsmouth, United Kingdom
yuting.hu@port.ac.uk

Shikun Zhou
School of Computing
University of Portsmouth
Portsmouth, United Kingdom
shikun.zhou@port.ac.uk

David Sanders
School of Computing
University of Portsmouth
Portsmouth, United Kingdom
david.sanders@port.ac.uk

Weicong Zhang
Zhejiang Wanli University
Zhejiang, China
zhangweicong@zwu.edu.cn

***Abstract*— The International Maritime Organization (IMO) enforced stricter sulphur abatement regulations since shipping emission has become one of the most major cause of the atmospheric pollution. Experts from the industry and academicians try to find the balanced solution among low-sulphur fuel, clean energy, and purposely fit scrubber by conventional statistical methods however failed to reach a satisfying conclusion. In addition, maritime datasets are usually massive, multi-source, and heterogeneous, it seems imperative for the maritime industry to adapt to the worldwide trend of intellectualisation and promote sustainable development.**

This work delineates and compares three main sulphur abatement solutions for ships through a thorough investigation of the current research state, and proposes a new framework based on a fusion model using modern big data and data mining algorithms. This work identifies and summarises major factors (with high impacts) in sulphur abatement solutions in the ocean shipping industry and integrate those high-level impacting factors to the proposed fusion model. The proposed framework can be optimised and utilised in determining suitable solutions for different ships, as well as shipping routes.

Keywords—sulphur abatement, maritime industry, big data, data mining, clustering algorithm

I. INTRODUCTION

Currently, the maritime industry is facing an arduous challenge that the International Maritime Organization (IMO) announced a new regulation that limits the sulphur emission of ships to 0.5% instead of 3.5% from 2020 [1]. Experts from the industry and academicians try to find the best solution among low-sulphur fuel, clean energy, and scrubber by conventional statistical methods. However, no general agreement has been made since the factors they considered are quite limited and each solution has both strengths and weaknesses. Therefore, utilising new data analysis techniques instead of traditional statistical methods in the maritime industry seems to be imperative, such as data mining techniques. Due to the current exponential growth of data, it is necessary to improve traditional data mining techniques to cope with a massive multi-source heterogeneous dataset, also known as big data. Clustering is an essential data mining technique that group samples into different clusters and the Gaussian Mixed Model (GMM) is one of the most popular and widely used clustering algorithms [2]. To fit GMM with a large dataset, this research will propose an optimised fusion model based on GMM. This model will be applied to evaluate sulphur abatement solutions to help the sustainable development of the shipping industry.

II. CURRENT DEVELOPMENT

A. New Regulations in Shipping

Shipping as the main transportation of international trade has increasingly developed with the boost of the world economy. According to the Review of Maritime Transport [3], a total of 42 million gross tons were added to global tonnage which indicated the expansion in ship supply capacity. The remarkable growth in world fleet capacity altered the shipping market balance and adjust freight rate, however, simultaneously causing severe environmental issues [4].

Attracting intensive international concerns, shipping emission has become one of the most significant reasons for atmospheric pollution [5]. The ship emissions that affect climate change mainly consist of carbon dioxide (CO_2), nitrogen oxides (NO_X), sulphur oxides (SO_X), and other pollutants [6]. Note that dioxides and trioxides can chemically react with exhaust gas or water in the air and generate caustic sulphurous acid and sulphuric acid which are the main substances of acid rain. The damage of acid rain is unaffordable for everyone, for instance, it can acidify soil and water, destroy vegetation and architecture, also affect the respiratory system of human beings and other animals [7].

Therefore, IMO which is a specialised agency of the United Nations (UN) and takes charge of making standards or regulations in the international shipping industry, decided at its 70th meeting of the Marine Environmental Protection Committee (MEPC) to limit the sulphur emission of ships to 0.5% instead of 3.5% from 2020 [1]. In the context of stricter emission regulations, how the existing fleet in the maritime industry can be adapted to the new emission regulations has become challenge for both shipowners and operators.

B. Shipping Dataset Analysis

According to IMO [8], there are mainly three alternatives to reduce ship sulphur emissions including switching to low-sulphur fuel oil, installing Exhaust Gas Cleaning System ("Scrubber"), and using eco-friendly fuel such as Liquefied Natural Gas (LNG). All of the three alternatives have their advantages and disadvantages. To help shipowners and operators to make a decision, related government departments, industry experts and scholars have made many efforts to evaluate the aforementioned sulphur abatement solutions.

Shifting to low-sulphur fuel might be the most convenient way for ships to comply with the new regulation because it only requires a minor modification on tanks and engines which is not a huge investment [9]. However, the operating costs will become a heavy burden for operators in the long run because low-sulphur fuel is much more expensive than heavy fuel oil [10].

Alternatively, installing a scrubber can allow ships to continue using cheap heavy fuel while the initial cost is considerably high as Antturi and colleagues mention that installing scrubbers may cost 1.5 million EUR per ship excluding pricey annual maintenance cost [11]. Considering the long payback period and opportunity cost during retrofitting, the market for scrubbers is still inconclusive [9]. Nevertheless, Solakivi and colleagues imply that scrubbers might be the mainstream method for sulphur abatement by conducting a descriptive analysis and logistic regression [12].

In terms of environmental performance, cleaner energy such as LNG is worth consideration because using LNG can significantly decrease emissions of NO_X, SO_X, and particulate matter (PM) [13]. LNG could be a good choice for LNG carriers or new buildings, however, it might not be suitable for other ships because of high up-front investments and the space-consuming volume which might reduce cargo capacity and cause loss of benefits [14].

There are also some comparative studies. Yang and colleagues develop a subjective generic methodology combining TOPSIS and AHP approaches to select NO_X and SO_X abatement techniques but they focus on fuel strategies in Sulphur Emission Control Areas (SECAs) [15]. Jiang, Kronbak & Christensen examine the costs and benefits of abatement approaches for ships and conclude that the price difference between low-sulphur fuel and LNG is determined the selection to a great extent and they suggest new ships to install scrubbers [16]. Abadie's team points out that the selection of abatement approaches depends on several factors, such as fuel price, operating area, ship type, ship age and sailing patterns [4]. They conduct an economic assessment of switching fuel and installing scrubbers under uncertainty and find that the remaining lifetime of a ship is a determinant factor. Lindstad, Rehn and Eskelan compare the three options as a function of ship types and operation patterns, their research indicates that scrubbers will be considered by large vessels while smaller vessels might choose to use fuels with less than 0.5% sulphur [17]. Wang and Peng point out that the size and shape difference affect shipping route choice and SO_2 emissions by sensitivity analysis [18]. Kim and Seo examine the actual response of Korean shipping companies by surveys and interviews and find that investment costs are the most vital factor, moreover, the companies' responses vary with their size [19].

C. *What Are the Problems*

Based on a rigorous review of the previous studies, it can be found that the ongoing intensive discussions surrounding sulphur abatement in the shipping industry mainly focus on separate solutions or comparisons between them, less on comprehensive analysis. Since each solution has its strength and weakness, there is not the best solution for the whole industry but the most suitable choice for each ship, each shipping route or each shipping company. However, there is too much data with too many factors involved and there is no one who combines and considers those key factors together to indicate the impact on ship sulphur abatement. This might because the digitalisation in the shipping industry is still ongoing and in shortage of data analysis technical personnel who are also professionals in international shipping. Therefore, an interdisciplinary study combining knowledge from both data science and international shipping transport is imperative. Moreover, traditional data mining algorithms also need optimisation to fit in the increasingly large dataset and apply in practice.

In general, traditionally sustainable shipping studies have followed contemporary regulatory developments, but they cannot adapt to stricter regulations announced by IMO. This research use data mining techniques to comprehensively investigate and identify major factors in sulphur abatement and help shipowners or operators to decide on the selection of abatement solutions. This research can encourage the shipping industry to save energy and reduce emissions by providing both beneficial and eco-friendly suggestions. Therefore, the proposed research is of great importance for the sustainable development of the shipping industry.

III. Identifying Major Impacting Factors of Sulphur Abatement Solutions

The Maritime industry is an integrated industry with many different complex processes. Due to its characteristics, maritime data is not only huge but also in various formats. There are many types of data produced during maritime activities, for instance, information on the bill of lading, navigation data, and other operational information. Fig.1 shows the main types of maritime data and they ought to be the domains in which big data and data mining technology should be developed.

Fig. 1. Types of Shipping Data

Based on the types of data that previous researches focused, major factors of sulphur abatement solutions were investigated and concluded as follows:

1) Economic factor (EF)

Subfactor EF1: one-off cost

Subfactor EF2: potential profits

Subfactor EF3: sustaining cost

Subfactor EF4: potential company growth

2) Operation factor (OF)

Subfactor OF1: types of ship

Subfactor OF2: shipping routes

Subfactor OF3: charter parties

Subfactor OF3: staffing especially crew members

Subfactor OF4: operation policy updates

3) Technology factor (TF)

Subfactor TF1: ship space organisation (usage)

Subfactor TF2: abatement equipment weights

Subfactor TF3: power consumption

Subfactor TF4: equipment lifetime

Subfactor TF5: maintenance technology

4) Abatement factor (AF)

Subfactor AF1: desulphurisation efficiency

Subfactor AF2: CO_X emission

Subfactor AF3: NO_X emission

Subfactor AF4: PM emission

Subfactor AF5: effluent and solid waste

These major factors will be evaluated in terms of data accessibility. Data might be collected using web crawler techniques through websites such as Automatic Identification System (AIS), Lloyd's List, Clarksons Shipping Intelligent Network and using any other approaches if necessary. The next stage of the work will be further exploration to each factor and classify identified factors into weighted categories. A weighing system will be proposed, and each factor will be assigned with different weights and those will be applied in proposed models later on.

IV. Developing Fusion Model in Decision Making

To discover useful information from a large dataset of historical observations, clustering techniques have been developed as an efficient method of grouping [20]. In the context of data mining, clustering means dividing a dataset into different clusters according to their different statistical behaviours to maximise the similarity of data in the same cluster and dissimilarity between clusters [21].

Al-Jubouri analyses the strength and weakness of representative clustering algorithms such as k-means algorithm, EM/GMM algorithm, normalized Laplacian spectral clustering and mean- shift clustering, he states that mixture models based on Gaussian distributions can explore clusters of various sizes and shapes [22]. However, the GMM algorithm might slow down when processing large data set and noise and outliers may prevent its efficiency [23]. Therefore, utilising GMM algorithm to generate clusters among big data is challenging for both academic and industry world.

GMM is a way of expressing K clusters of a data set. Each cluster is treated as a multivariate Gaussian distribution with mean vectors μ and covariance matrices R as parameters [2]. Each distribution is a component of the mixture model and k is known as the order of the mixture model. Given a data set $X = \{x_1, x_2, ..., x_N\}$ of d dimensions, the GMM is represented as $\Theta = \{\theta_1, \theta_2, ..., \theta_K\}$ where $\theta_k = (\mu_k, R_k)$ $(1 \leqslant k \leqslant K)$. If $p(x_n|\theta_k)$ represents the probability that data object x_n is drawn from the k^{th} distribution θ_k, and a_k represents the probability that the k^{th} distribution is chosen, then:

$$p(x|\Theta) = \sum_{k=1}^{K} a_k p(x_n|\theta_k), \qquad \sum_{k=1}^{K} a_k = 1$$

For GMM, $p(x_k|\theta_k)$ is often taken as the probability density function for Gaussian distribution:

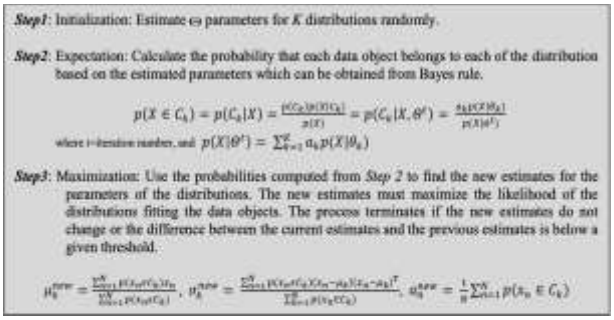
$$p(\theta_k) = \frac{1}{(2\pi)^{d/2}} |R_k|^{1/2} exp\, \{-\frac{1}{2}(x_n - \mu_k)^t R_k^{-1}(x_n - \mu_k)\}$$

Assuming that each data object is drawn independently, the probability of obtaining the whole data set is therefore:

$$p(X|\Theta) = \Pi_{n=1}^{N} \sum_{k=1}^{K} a_k p(x_n|\theta_k)$$

The logarithm of the function above is known as the log-likelihood function. The objective of GMM clustering is to estimate the parameters in Θ concerning X such that the function value is maximized, indicating that the data set is the most likely result modelled by the GMM.

The Expectation-Maximization (EM/GMM) algorithm is used to find the fittest GMM for a data set [2]. The basic steps of the algorithm are indicated in Figure 2.


Step1: Initialization: Estimate Θ parameters for *K* distributions randomly.

Step2: Expectation: Calculate the probability that each data object belongs to each of the distribution based on the estimated parameters which can be obtained from Bayes rule.

$$p(X \in C_k) = p(C_k|X) = \frac{p(C_k)p(X|C_k)}{p(X)} = p(C_k|X, \theta^i) = \frac{a_k p(X|\theta_k)}{p(X|\theta^i)}$$

where i=iteration number, and $p(X|\theta^i) = \sum_{k=1}^{K} a_k p(X|\theta_k)$

Step3: Maximization: Use the probabilities computed from *Step 2* to find the new estimates for the parameters of the distributions. The new estimates must maximize the likelihood of the distributions fitting the data objects. The process terminates if the new estimates do not change or the difference between the current estimates and the previous estimates is below a given threshold.

$$\mu_k^{new} = \frac{\sum_{n=1}^{N} p(x_n \in C_k) x_n}{\sum_{n=1}^{N} p(x_n \in C_k)}, \; R_k^{new} = \frac{\sum_{n=1}^{N} p(x_n \in C_k)(x_n - \mu_k)(x_n - \mu_k)^T}{\sum_{n=1}^{N} p(x_n \in C_k)}, \; a_k^{new} = \frac{1}{N}\sum_{n=1}^{N} p(x_n \in C_k)$$

Fig. 2. Basic EM/GMM algorithm [24]

Compared with other clustering algorithms such as k-means, the GMM clustering algorithm assumes that features of samples obey Gaussian distribution, which is more suitable for central limit theory under the context of large samples. Moreover, the final result of the GMM algorithm will be the probability of each sample belonging to different clusters, which is more reasonable than k-means. Therefore, the GMM algorithm is identified as a start to combine selected key factors, and then it will be changed to fit large datasets and the specific problem.

To apply the proposed fusion model towards sulphur abatement solutions, we assume that each factor obeys the Gaussian distribution. The four major factors (with high impacts) can be seemed as four dimensions. Because there are only four potential choices for each ship, which respectively are switching to low-sulphur fuel oil, installing Scrubber and using eco-friendly fuel such as LNG installing scrubber, or scrapping the ship. There should be four components of the proposed GMM model, which means K is four in this circumstance.

Here is a roughly designed framework of applying the fusion model.

1) Pre-process (feature scaling and normalisation) and input raw data

2) Calculate high-impact factors with weights and subfactors.

3) Compute the Initial clustering and estimate parameters.

4) Expectation step: calculate the latent variable (the probability that each ship belongs to each distribution) based on initial parameters.

5) Maximisation step: use the latent variable to estimate new parameters.

6) Iterate expectation step and maximisation step until convergence.

7) Output parameters and results of the belongings of each ship.

This framework will be trained and tested to be optimised so that the proposed fusion model based on GMM can be applied to give a reasonable and sensible output which indicates the selection of sulphur abatement solutions.

V. Validation and Examples

In order to verify the effectiveness of the proposed model, experimental verification was performed on Matlab.

An example dataset is shown in Figure 3. It demonstrates that the relationships between the consumption of heavy and low sulphur fuel oil consumptions and voyages, dead weight, sailing mileage, sailing time and berth time. It is no surprise that the usage of heavy fuel oil is proportional to the sailing time and mileage.

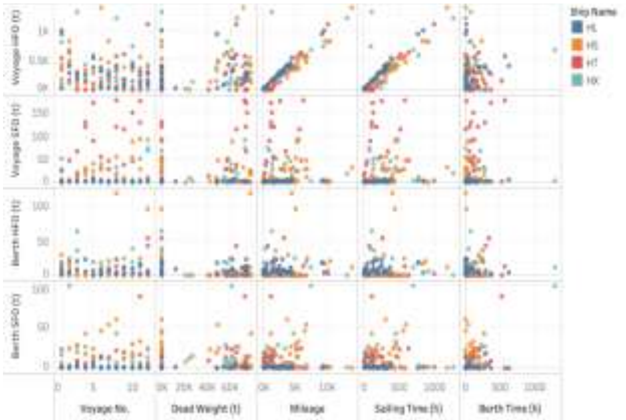


Fig. 3. Scatter plot matrix of fuel consumption for four randomly selected ships, measured between 2016 and 2018.

The aim of the experiment is to use data mining algorithm to help with the determination of sulphur abatement solutions. Applying the clustering algorithm into the dataset, it can be found that four ships are assigned to three clusters. By comparing the clustering result and the real decision made by ship owners, regular patterns can be discovered.

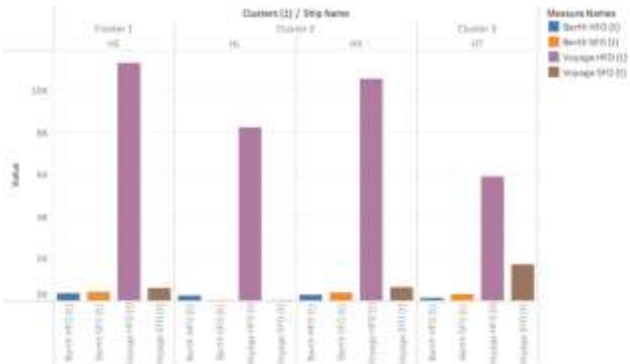


Fig. 4. Histogram of four ships in three clusters according to fuel consumption.

The three ship sulphur abatement solutions have their own advantages and disadvantages. It might cost huge economic price and manpower resouces for human judgement. But with the help of model and algotithms, we can easily get reliable suggestion on wheter this ship should choose low sulphur fuel oil, install a scruber or change to LNG.

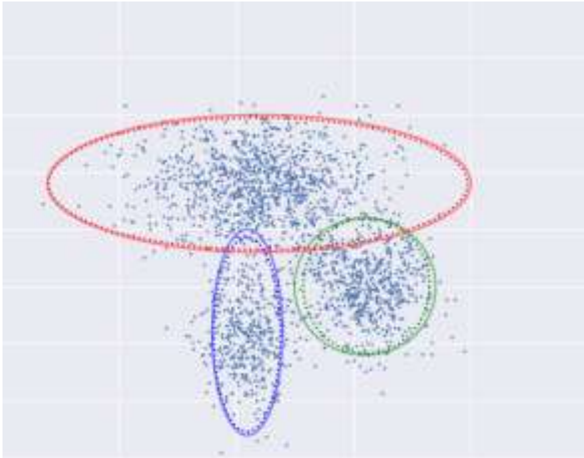


Fig. 5. Example of clustering result

Based on traditional statistic methods and optimised by data mining algorithm , this new model improved the accuracy of clustering result and saved resources. However, comparing to the original intention of this research, there are still many limitations. For example, we did not take the variation of fuel price into account, which needs further works.

VI. Conclusions and Future Work

This work aims to propose and apply a fusion model based on data mining algorithms to evaluate sulphur abatement solutions in the maritime industry because the maritime industry is facing the challenge of stricter regulations for sulphur abatement and there is no comprehensive evaluation or analysis of solutions. We propose to utilise data mining techniques to handle the problem while traditional clustering algorithms such as Gaussian Mixed Model need to be optimised to apply in practice. In expectation, a new fusion-based methodology can be developed to combine and integrate identified key factors with high impact on the selection of sulphur abatement solutions. So that the proposed research can contribute to not only the practical application of data mining techniques but also the sustainable development of the maritime industry.

For future work, firstly the proposed fusion model will be completed in detail and optimised for utilisation. Then, more maritime data shall be collected to train and test the model. Once we obtained enough data, they can be clustered into four groups which respectively are switching to low-sulphur fuel oil, installing Scrubber and using eco-friendly fuel such as LNG installing scrubber, or scrapping the ship. When the involved dataset becomes mass, we will try to conduct this work on some big data platforms such as Hadoop, Hive, and Spark and figure out how to adapt it.

Acknowledgment

We thank the University of Portsmouth for its help and support to this work.

REFERENCES

[1] IMO, "Marine Environment Protection Committee (MEPC) 70th session," London, 2016. [Online]. Available: http://www.imo.org/en/MediaCentre/MeetingSummaries/MEPC/Pages/MEPC-70th-session.aspx.

[2] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning (Corrected 12th printing)*. New York: Springer, 2017.

[3] UNCTAD, *Reveiw of Maritime Transport*. New York: United Nations Publications, 2018.

[4] L. M. Abadie, N. Goicoechea, and I. Galarraga, "Adapting the shipping sector to stricter emissions regulations: Fuel switching or installing a scrubber?," *Transp. Res. Part D Transp. Environ.*, vol. 57, no. October, pp. 237–250, 2017, doi: 10.1016/j.trd.2017.09.017.

[5] Ø. Endresen *et al.*, "Emission from international sea transportation and environmental impact," *J. Geophys. Res. D Atmos.*, vol. 108, no. 17, 2003, doi: 10.1029/2002jd002898.

[6] I. M. Hassellöv, D. R. Turner, A. Lauer, and J. J. Corbett, "Shipping contributes to ocean acidification," *Geophys. Res. Lett.*, vol. 40, no. 11, pp. 2731–2736, 2013, doi: 10.1002/grl.50521.

[7] A. Singh and M. Agrawal, "Acid rain and its ecological consequences," *J. Environ. Biol.*, vol. 29, pp. 15–24, 2008.

[8] IMO, "IMO 2020 - A Breath of Fresh Air," 2019. https://imo-newsroom.prgloo.com/resources/imo-2020-a-breath-of-fresh-air.

[9] R. Bergqvist, M. Turesson, and A. Weddmark, "Sulphur emission control areas and transport strategies -the case of Sweden and the forest industry," *Eur. Transp. Res. Rev.*, vol. 7, no. 2, 2015, doi: 10.1007/s12544-015-0161-9.

[10] Y. Gu and S. W. Wallace, "Scrubber: A potentially overestimated compliance method for the Emission Control Areas: The importance of involving a ship's sailing pattern in the evaluation," *Transp. Res. Part D Transp. Environ.*, vol. 55, no. 2017, pp. 51–66, 2017, doi: 10.1016/j.trd.2017.06.024.

[11] J. Antturi *et al.*, "Costs and benefits of low-sulphur fuel standard for Baltic Sea shipping," *J. Environ. Manage.*, vol. 184, pp. 431–440, 2016, doi: 10.1016/j.jenvman.2016.09.064.

[12] T. Solakivi, S. Laari, T. Kiiski, J. Töyli, and L. Ojala, "How shipowners have adapted to sulphur regulations – Evidence from Finnish seaborne trade," *Case Stud. Transp. Policy*, vol. 7, no. 2, pp. 338–345, 2019, doi: 10.1016/j.cstp.2019.03.010.

[13] M. Acciaro, "Real option analysis for environmental compliance: LNG and emission control areas," *Transp. Res. Part D Transp. Environ.*, vol. 28, no. 2014, pp. 41–50, 2014, doi: 10.1016/j.trd.2013.12.007.

[14] H. Ma, K. Steernberg, X. Riera-Palou, and N. Tait, "Well-to-wake energy and greenhouse gas analysis of SOX abatement options for the marine industry," *Transp. Res. Part D Transp. Environ.*, vol. 17, no. 4, pp. 301–308, 2012, doi: 10.1016/j.trd.2012.01.005.

[15] Z. L. Yang *et al.*, "Selection of techniques for reducing shipping NOx and SOx emissions," *Transp. Res. Part D Transp. Environ.*, vol. 17, no. 6, pp. 478–486, 2012, doi: 10.1016/j.trd.2012.05.010.

[16] L. Jiang, J. Kronbak, and L. P. Christensen, "The costs and benefits of sulphur reduction measures: Sulphur scrubbers versus marine gas oil," *Transp. Res. Part D Transp. Environ.*, vol. 28, no. 2014, pp. 19–27, 2014, doi: 10.1016/j.trd.2013.12.005.

[17] H. E. Lindstad, C. F. Rehn, and G. S. Eskeland, "Sulphur abatement globally in maritime shipping," *Transp. Res. Part D Transp. Environ.*, vol. 57, pp. 303–313, 2017, doi: 10.1016/j.trd.2017.09.028.

[18] S. Wang and C. Peng, "Model and analysis of the effect of China's potential domestic emission control area with 0.1% sulphur limit," *Marit. Bus. Rev.*, vol. 4, no. 3, pp. 298–309, 2019, doi: 10.1108/MABR-03-2019-0012.

[19] A. R. Kim and Y. J. Seo, "The reduction of SOx emissions in the shipping industry: The case of Korean companies," *Mar. Policy*, vol. 100, no. September 2018, pp. 98–106, 2019, doi: 10.1016/j.marpol.2018.11.024.

[20] M. Sun, "Big data analytics in power systems," Imperial College London, 2016.

[21] B. Everitt, *Cluster Analysis*, 2nd ed. New York: Halsted Press, 1980.

[22] H. Al-jubouri, "Multi Evidence Fusion Scheme for Content-Based Image Retrieval by Clustering Localised Colour and Texture Features BY," no. June, 2015.

[23] P.-N. Tan, M. Steinbach, and V. Kumar, *Introduction to Data Mining*. Boston: Pearson, 2006.

[24] H. Du, *Data mining techniques applications: an introduction*. Hampshire: Cengage learning, 2010.