International Journal on Advanced Science Engineering Information Technology

Vol.14 (2024) No. 6
ISSN: 2088-5334

Data Analysis on Factors Influencing Oceanic Plastic Pollution

Meng-Chew Leow*, Chan-Jin Loo, Lee-Yeng Ong

Faculty of Information & Science Technology, Multimedia University, Melaka, Malaysia

**Corresponding author: mcleow@mmu.edu.my*

***Abstract*—Oceanic pollution is a serious environmental problem that threatens the lives of marine life nowadays. One of the main reasons is that marine life accidentally eats the litter that falls into the ocean. Other pollutants included microplastics, discarded chemicals, animal carcasses, etc. According to statistics, millions of animals die annually from accidentally eating plastic. This catastrophe is caused by the higher use of plastics due to the convenience it brings to people. However, they are often thrown away after a single use, and the plastic waste ends up in the ocean due to improper disposal. Therefore, the main aim of this paper is to understand the incident of oceanic pollution based on some mismanaged plastic (MMP) waste variables. The data analytic method is used to understand the relationships among MMP variables. Visualizations of the analysis results, such as linear scatter plots and some analysis tables, provide valuable insights into ocean pollution based on MMP waste variables. This study helps develop informed policies for practical oceanic plastic pollution mitigation efforts. To address ocean pollution effectively, we must develop sustainable practices that enhance the health of our seas. Global organizations must also take strong measures to combat the increasing threat of plastic pollution. By implementing targeted action plans, we can protect marine ecosystems and ensure a cleaner future for our oceans.**

***Keywords*—Linear regression; multiple regression; mismanaged plastic waste; ocean pollution; plastic pollution**

Manuscript received 8 Dec. 2023; revised 9 Feb. 2024; accepted 15 Sep. 2024. Date of publication 31 Dec. 2024.

I. INTRODUCTION

Ocean plastic pollution is a critical environmental issue that has existed for a long time [1], [2], [3], [4]. The problem affected marine ecosystems [5], wildlife [6], and human health [7], [8], [9], [10]. As the accumulation of plastic debris in the ocean poses a multifaceted threat, comprehensive research is needed to understand the solutions [11], [12], [13], [14]. However, these environmental problems are often man-made [3]. People buy or use disposable plastic items, including plastic bags, plastic bottles, food wrappers, etc. Still, they do not dispose of them, and instead, they adequately litter them randomly. This ultimately results in this littered waste reaching waterways and the ocean. Therefore, this project is mainly focused on exploring the correlation between the variables of mismanaged plastic waste and ocean pollution. Understanding the correlation requires using some analytical methods to present the visualization and results of the analysis to find out the influencing factors [15], [16], [17].

This paper aims to explore the correlations between the variables of mismanaged plastic (MMP) waste. One of the problems is determining which two variables of MMP waste will influence each other by performing the linear regression models. This is because the variables that interact with each other can provide an enjoyable analytical process to get a helpful result. Another goal of this paper is to develop a multiple regression model for prediction to capture the complexity of deeper relationships. The problem statement to achieve the goal is to determine which independent variables are influential factors in the dependent variable. The multiple regression model will be implemented to remove the non-significant independent variables based on their p-value and rerun the model to evaluate an accurate prediction result. The first two objectives will be achieved after implementing the linear and multiple regression models. Finally, the last purpose of this paper is to assess the impact of mismanaged plastic waste in the ocean based on analytical methods. The problem statement for this goal is based on completing the first two goals because we need to get their results to identify the impacts of MMP waste in the ocean.

A. Ocean Pollution

Billions of pounds of trash and other pollutants pollute the ocean every year. Ocean debris has become a persistent pollution problem that affects the entire ocean and Great Lakes. The pollutants included microplastics, discarded chemicals, animal carcasses, etc. Where do the pollutants come from? Most pollutants come from human activities such as point- and non-point-source pollution. A point source of pollution is a single and easily identifiable source (e.g. [18]). It comes from a single specific place, such as a smokestack, drain, or water pipe. Pollution from damaged or malfunctioning industries or water treatment plants is classified as a point source of pollution. Some industrial wastewater is often discharged into the ocean in this way. Fortunately, point source pollution incidents are rare and frequently have significant effects. Additionally, point source pollution is also relatively easy to control and restricted by relevant units. For example, restricting the flow of untreated wastewater into the ocean. Therefore, this pollution is not a big threat to polluting the ocean.

On the contrary, non-point source pollution is the opposite of point source pollution. Its pollutants discharge range is wider than point source pollution (e.g. [19], [20]). Non-point source pollution can include toxic chemicals from urban areas, septic tanks, excess fertilizers from agricultural lands, and so on. How was the non-point source pollution formed? Non-point source pollution is caused by rainfall, land runoff, or melting snow moving over the ground. Along the path of the water flow, it takes away the things passing by, such as residual oil droplets on the road, garbage, and blockage items in ditches, and eventually, these pollutants will drift into the ocean and river areas. It can also be known as a natural concept of pollution that people cannot control.

B. Plastic Pollution

Plastics are made from fossil fuels and have existed in this world for over a thousand years. Plastic provides people with a lot of convenience but also leads to a throw-away culture. For example, overuse of disposable products such as disposable straws, plastic cups, plastic bags, and so on. Today, disposable plastics account for 40% of annual plastic production. This includes plastic bags and food wrappers. The rapid growth of disposable plastic product production exceeds the world's ability to deal with them. However, the service life of these items is less than a day or even a few minutes. But they may remain in the environment for hundreds of years and contribute to the world's plastic pollution (e.g. [21], [22]).

Plastic pollution is a global problem that has been a concern worldwide [23]. Although oceans cover more than 70% of the Earth's surface, it is hard to imagine the big amount of plastic being disposed into the sea every day. Every day, a volume of plastic equivalent to that carried by 2,000 garbage trucks is disposed into the oceans, rivers, and lakes worldwide while an estimated 19–23 million tons of plastic waste are disposed into lakes, rivers, aquatic ecosystems, and seas every year (e.g. [24], [25]). According to statistics, millions of animals die every year due to accidentally eating plastic. These animals include nearly 700 species that have been affected by plastic, including endangered ones. Most animals accidentally eat plastic because they are starving. But not all animals are like that case, especially marine organisms. Marine organisms are often strangled by discarded fishing gear or discarded large plastic tools. All in all, plastic pollution is one of the most pressing environmental issues [26], [27], [28], [29]. It has seriously affected the living environment of animals and threatened their lives. In the news on 08 March 2023, "More than 170 trillion plastic particles found in the ocean as pollution reaches 'unprecedented' levels," by Laura Paddison was published in Cable News Network (CNN).

A new study shows that the world's oceans are polluted by an estimated 171 trillion plastic particles, which means that around 2.3 million tons are gathered. This study is analyzed by a team of international scientists. They collected global data from 12,000 sampling points which are in the Atlantic, Pacific, and Indian Oceans and the Mediterranean Sea between 1979 and 2019. Based on their study, they found that ocean plastic pollution unprecedentedly increased at a rapid rate since 2005. Lisa Erdle, who was the director of research and innovation at the 5 Gyres Institute and an author of the report, told CNN that the rate of ocean plastic pollution was higher than previously estimated. Therefore, the study found that the rate of plastics entering the oceans will increase by around 2.6 times from now and until 2040. This means that the 2.3 million tons of plastic particles will multiply by 2.6 times, which is a total of around 5.98 million tons of plastic particles will be in the ocean by 2040.

What was the main reason for the rising rate of ocean plastic pollution? The main source of this reason is still traced down to human behavior. Nowadays, people commonly use plastic products and throw them away randomly. Overuse of plastic products causes the plastic to be mass-produced without proper disposal. Once the supplies are sufficient, people will repeat the same mistakes. In this way, the plastic discarded increased rapidly. The huge amount of discarded plastic that comes from land will be transported out to the ocean by rain, wind, and drains eventually. The plastic does not decompose once it enters the ocean, becoming plastic debris. This is why dealing with the oceanic plastic waste is a difficult matter. In addition, the plastic debris may be ingested by marine life, affecting the marine ecosystem and causing environmental disasters. Most plastic raw materials are fossil fuels, which also contribute to catastrophic climate problems. This is because the lifecycle for producing fossil fuels until its disposal is causing planet-heating pollution.

All in all, disposing of plastic waste pollution in the ocean is not easy. Oceans occupy 71% of the earth and are complex places with dramatic weather changes. The worsening oceanic pollution problem will need urgent attention from the world population to really work out an effective solution. Oceanic space conservation needs to appeal to humans by reducing the use of plastic products and managing garbage waste properly.

II. Materials and Method

A. Data Source and Variables

The data sources and variables used in this paper, according to the country data, include the mismanaged plastic waste to ocean per capita, mismanaged plastic waste per capita, mismanaged waste emitted to the ocean, amount of plastic emitted to the ocean, and total mismanaged plastic waste. These data sources of MMP variables were found from "Kaggle" and "Our World in Data". After that, combine them based on the same year (2019) and gather with specific countries. In addition, their units are converted to kilograms (kg) to implement the analysis. To get more accurate results, researchers preprocess the data by handling missing data, correcting the inconsistencies, and ensuring data are not duplicated so that it can prevent skewing of results.

B. Regression Techniques

This paper implemented two types of regression techniques: linear regression and multiple regression. The linear regression will be implemented using the dependent variable (the outcome or response variable was denoted "Y") and the independent variable (the variable that was manipulated or predicted was denoted "X"). This paper will conduct three types of linear regression analysis between two variables among the data sources above.

By using linear regression analysis, we can find out the correlation relationship between two variables of data sources. Afterward, the analysis will display the linear equation (Y=mX+C) and R-squared result with a scatter plot. The linear equation can be used for the prediction of the related result based on the correlation between two variables. Then, the hypothesis will be explored when referring to the linear equation. The R-squared, or the coefficient of determination, is used to determine the proportion of variance between the dependent variable that can be explained by the independent variable. In other words, it provides information about the goodness of fit of the model. The R-squared values range from 0 to 1. To identify its result, the higher the R-squared is displayed, the better the model fits the data used. Besides, the r value (Pearson correlation coefficient) in the scatter plot also will be conducted. The number of r values is between -1 and 1, which is used to offer a quantitative measure between two variables in the scatter plot.

a. When the r value is positive correlation ($r>0$), means both variables in the scatter plot tend to increase at the same time.
b. When the r value is a negative correlation ($r<0$), it means one variable is increasing, and another one is decreasing.
c. When the r value is equal to 0 ($r=0$), means there is no linear relationship between both variables in the scatter plot.

Furthermore, the hypothesis of linear regression can also be implemented, which is the variable of the scatter plot will influence another variable. There are some hypotheses of linear regression analysis such as:

a. When coefficient (m) is positive, then the independent variable (X) increases and the dependent variable (Y) will increase.
b. When coefficient (m) is negative, then the independent variable (X) increases and the dependent variable (Y) will decrease.
c. When coefficient (m) is positive, then the independent variable (X) decreases, and the dependent variable (Y) will decrease.
d. When coefficient (m) is negative, then the independent variable (X) is decreasing, and the dependent variable (Y) will increase.

Another regression technique is multiple regression analysis based on the five variables of the data source. Multiple regression is implemented by using one dependent variable (Y) and two or more independent variables (X). The dependent variable represents a variable that needs to be predicted or explained based on the independent variables (predictors or explanatory variables).

By using multiple regression analysis, we can understand how changes in the independent variables are related to changes in the dependent variable. After the summary output of multiple regression analysis, the summary output of multiple regression analysis will show the regression statistics table and ANOVA table and interpret regression coefficients. In the multiple regression statistics table, the table will have five results values such as multiple R, R squared, adjusted R square, standard error of the regression and observations. Multiple R is the positive square root of R squared. It is the correlation coefficient that determines the strong or no linear relationship at all. A positive relationship is when the value of multiple R is 1, and the zero value of multiple R means no relationship at all. R squared is the coefficient of determination and is the square of the multiple R-values. For example, the value of R squared is 0.70, which means that 70% of the variation of the dependent variable (Y) can be explained by independent variables (X). The higher value of R squared value can make the efficiency of prediction and explain more variation of the model. On the contrary, a lower R squared value indicates the difficulty of the variation between dependent and independent variables because the model didn't fit the data. It is used to determine how many points fall on the regression line and represents the power of the model. Adjusted R squared is a modified version of R-squared that is used to adjust the non-significant predictors in the model. The standard error of regression is the accuracy of regression coefficient measurement. Observations means the number of data that is used for analysis.

Next, the second part will show the ANOVA table of multiple regression analysis output. The results of the ANOVA table show the regression and residual value of degree of freedom (df), sum of squares (SS), mean squared error (MS), F test for null hypothesis (F), and significance associated P-value (Significance F).

The last part of the output of the multiple regression analysis is the interpreted regression coefficients table. The table shows the result of the coefficient, standard error, T statistic, P-value, lower 95% value, and upper 95% value. For this paper, the p-value is the most important to examine so that we can proceed more deeply. In the interpreted regression coefficients table, the standard error provides the estimated standard deviation by which the coefficient varies across different cases. T-statistic is equal to the coefficient value divided by the standard error value. The larger value of the T-statistic will be when the larger coefficient with respect to the standard error. The important value that needs to be examined is the p-value. The P-value is determined by the comparison between t-statistic and t-distribution. There are two situations that will be identified when examining the p-value. When the p-value is less than the significant level (often 0.05), it indicates that the model has a significant linear relationship between the dependent and independent variables. On the contrary, when a p-value is more than 0.05, it means that there is not sufficient evidence between dependent and independent variables.

Afterward, we need to look at the Significance F as less than 0.05 so that the overall regression model is statistically significant. Then, make sure the p-value of the independent variables is less than the chosen significance level (0.05). The significance level is the probability of rejecting the null hypothesis. There are some specific guidelines to indicate the p-value such as:

a. when p-value ≥ 0.1, it means the data consistent with the null hypothesis.
b. when $0.05 \leq$ p-value < 0.1, it means low evidence against the null hypothesis.
c. when $0.01 \leq$ p-value < 0.5, it means moderate evidence against the null hypothesis.
d. when $0.0001 \leq$ p-value < 0.01, means strong evidence against the null hypothesis.
e. when p-value < 0.001, it means powerful evidence against the null hypothesis.

However, in this paper, the p-value needs to be less than the significant level (0.05) so that it can reject the null hypothesis. If one of the p-values of independent variables is more significant than 0.05, we need to remove the variable and rerun the multiple regression model. When all p-values of independent variables are less than 0.05, we can identify whether the model is a null hypothesis (H0) or an alternative hypothesis (H1). The null hypothesis means no relationship exists between the independent variables and the dependent variable. Mathematically, it can be expressed as: (H0: $\beta_1 = \beta_2 = ... = \beta_k = 0$, β represents the independent variable in the model). On the contrary, the alternative hypothesis (H1) means that at least one of the independent variables has a significant relationship with the dependent variable (H1: At least one $\beta_e \neq 0$).

A prediction macro can be successfully calculated when all coefficients are statistically significant at a 5% significance level. The prediction macro predicts the value of the dependent variable based on the independent variables. The formulae are calculated based on the coefficients of intercept plus the coefficients of each independent variable multiplied by a value. Then, the predicted number of dependent variables is successfully calculated. The analysis flow is illustrated in Fig. 1.

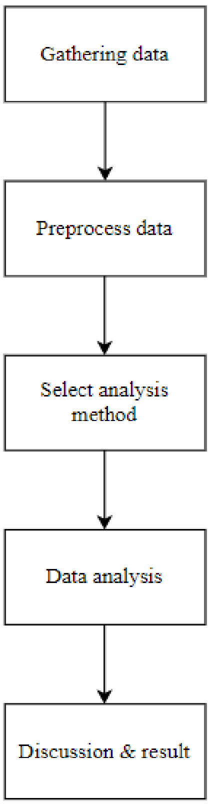

Fig. 1 Analysis flowchart

Researchers need to gather data from different datasets before performing the analysis work. This is because the datasets were found from different dataset online platforms such as “Kaggle” and “Our World in Data.” Therefore, the different data sources need to be gathered into a dataset so that it is easy to analyze later. The similarity of these data is the same year and based on countries. Firstly, researchers combined them based on the same year (2019) and gathered with specific countries.

After completing gathering and combining the datasets, researchers need to preprocess the dataset. To get more accurate results, researchers preprocess the data by handling missing data, correcting the inconsistencies, and ensuring data are not duplicated so that it can prevent skewing of results. When the data is combined with the same country, the missing and duplicate data will be found. This is because, at most, two columns of different data are used only during the combination process. Therefore, the problem of missing data and duplicate data can be easily found.

Then, continue to combine other columns of data through this process and remove the problems. After successfully handling the problems, the dataset needs to be converted to the same numerical format as the kilogram’s unit. Some of the actual data are in kilograms and some are in grams. Then, the data with the gram units needs to be divided by 1000kg to get the kilogram unit. The data unit consistency is to ensure the easily implemented data analysis and obtain the accuracy of analysis results. Thus, the important part of data preprocessing is to prevent inaccurate analysis results because it will affect the performance of the analysis model and lead to biased results.

When the data preprocessing is completed, data analysis can be implemented. There are two types of analysis methods such as linear regression and multiple regression analysis. The important part is to consider the suitable two variables of data to investigate the linear regression analysis. This is because we need to get a standard result of the R-squared value between each variable so that we can have a strong relationship. Therefore, need to attempt different pairings of two variables among the dataset to get the most effective results and implement it. For the multiple regression analysis, we need to identify a dependent variable, and it is also a predictive value of the model. Then, we need to be consistent with many factors to get the best-fit model. Related factors such as the p-value of the independent variables and the R-squared value of the model result. The p-value of each independent variable needs to be less than the significance level (0.05) and the R-squared value needs to be above 0.7 to get a high level of correlation.

During the data analysis process, some problems were encountered. For example, the R-squared result of linear scatter plots and multiple regression obtain a very low value. It means that the linear scatter plot does not align with a straight line. In other words, there is no relationship between the independent variable and the dependent variable of the plot. Due to the weak relationship of linear scatter plot obtained, the independent variables of multiple regression are not explaining much of the variability in the dependent variable. Therefore, there is a hidden process which is needed to find more datasets that can be suitable to match with the dataset that used. Afterward, continue the process of gathering the data, preprocessing data, and test-running the draft analysis. This process is to ensure that the validity of the analysis results is obtained.

After the data analysis results of both linear regression and multiple regression are obtained, discussion and results will be conducted. The linear regression will be explained based on the linear equation and R-squared through the linear scatter plots. In addition, the results of multiple regression will be explained using some statistic tables. The tables include multiple regression statistic tables, ANOVA tables, and interpret regression coefficients tables. Each table is an important factor in explaining the results.

III. Results and Discussion

A. Linear Regression Results

Fig. 2 shows the correlation between the amount of plastic emitted to the ocean (independent variable) and mismanaged waste emitted to the ocean (dependent variable). The purpose of developing this linear regression is to determine whether the mismanaged waste emitted to the ocean is influenced by the amount of plastic emitted. The results show a higher R-squared value of 0.7993, which means that there is a strong relationship between these two variables. Another more precise result can be the result of the R-squared value, which shows that 79.93% of the variability in mismanaged waste emitted to the ocean can be explained by the amount of plastic emitted to the ocean. Therefore, it can be hypothesized that the increasing amount of plastic emitted to the ocean will lead to the increase of mismanaged waste emitted to the ocean. Although other factors will affect the amount of plastic emitted into the ocean, such as plastic litter from sewers, shipping activities, and recreational activities, the major contributor is mismanaged waste. This is because humans throwing garbage daily will lead to much garbage being disposed of improperly every time.

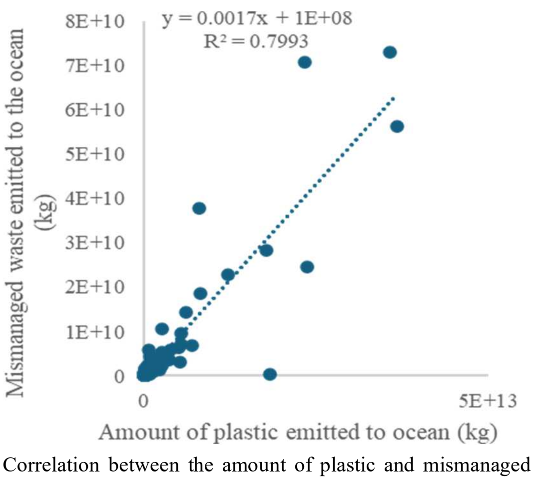

Fig. 2 Correlation between the amount of plastic and mismanaged waste emitted to the ocean

Fig. 3 shows the correlation between MMP waste (independent variable) and MMP waste to ocean (dependent variable) per capita. The purpose of investigating this linear regression is to evaluate the impact of MMP waste per capita on the amount of plastic waste in the ocean. As a result, the R-squared value of the linear regression is 0.5887, which indicates that approximately 58.87% of the variability in the MMP waste to the ocean can be explained by the MMP waste per capita. In other words, mismanagement of plastic waste by human behavior has seriously affected ocean plastic pollution. Besides, the linear equation ($y=0.0641x-0.1365$) means that if mismanaged plastic waste per capita increases by 1kg, then mismanaged plastic waste to the ocean per capita increases by 0.0641kg. This indicated that increasing mismanaged plastic waste per capita will lead to an increase of mismanaged plastic waste in the ocean per capita. Therefore, people always dispose of garbage casually in their daily lives, which is a very serious problem that causes ocean plastic pollution.

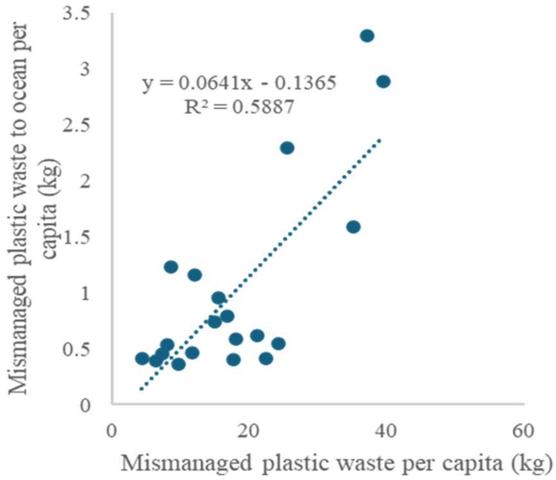

Fig. 3 Correlation between MMP waste and MMP waste to the ocean (per capita)

Fig. 4 shows the correlation between total MMP waste (independent variable) and mismanaged waste emitted to the ocean (dependent variable). Investigating this linear regression aims to support the hypothesis of Figure 3 result to provide stronger evidence. An R-squared value of 0.6934, as shown above, indicates 69.34% of the variability in mismanaged waste emitted to the ocean can be explained by the total MMP waste. Based on the result of the linear equation ($y=8E\text{-}06x+2E\text{+}09$), a hypothesis can be drawn as total mismanaged plastic waste increases by 1kg. Mismanaged waste emitted to the ocean increases by 0.000008kg (8E-06). Although this hypothesis is not the weightiest factor that causes ocean pollution, the sea still gets polluted due to the large population on Earth.

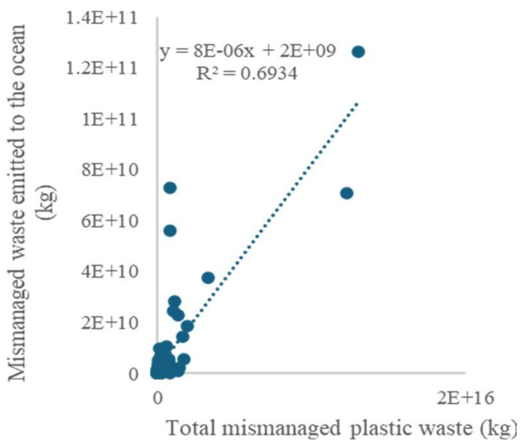

Fig. 4 Correlation between total MMP Waste and mismanaged waste emitted to the ocean

B. Multiple Regression Results

Table I is the first model of multiple regression analysis based on the data sources used in this paper. The independent variables in this multiple regression analysis included mismanaged plastic waste to the ocean per capita, mismanaged plastic waste per capita, mismanaged waste emitted to the ocean, and the amount of plastic emitted to the ocean. The dependent variable used is total mismanaged plastic waste because it enabled the achievement of an interesting outcome that can be understood or predicted based on other factors. The result of the first model multiple regression analysis shows a strong R-Square value of 0.7966 based on 99 observations.

TABLE I
MULTIPLE REGRESSION STATISTICS (MODEL 1)

Regression Statistics	
Multiple R	0.892526
R Square	0.796602
Adjusted R Square	0.787947
Standard Error	8.54E+14
Observations	99

Table II shows the ANOVA table results with a significant F value less than the significance level (0.05). This means that the overall model rejected the null hypothesis and concluded that the model's results were statistically significant.

TABLE II
ANOVA TABLE (MODEL 1)

ANOVA	**df**	**SS**	**MS**	**F**	**Sig. F**	
Regression	4	2.687E+32	6.7E+31	92.037015	1.1942E-31	<0.05
Residual	94	6.861E+31	7.3E+29			
Total	98	3.373E+32				

However, when checking the p-values of all independent variables in the regression output table of Table III, we can see that the p-value of the independent variable (b2) is higher than 0.05. In this case, we need to remove the independent variable with a higher than 0.05 p-value and rerun the multiple regression analysis again so that we can get enough evidence to reject the null hypothesis.

TABLE III
INTERPRET REGRESSION COEFFICIENTS (MODEL 1)

		Coefficients	**Standard Error**	**t Stat**	**P-value**	**Lower 95%**	**Upper 95%**	**Lower 95.0%**	**Upper 95.0%**
b 0	Intercept	1.21E+14	1.25E+14	0.97	0.334093	-1.26E+14	3.68E+14	-1.26E+14	3.68E+14
b 1	Mismanaged plastic waste to ocean per capita person (kg per year)	-9.53E+14	2.39E+14	-3.98	0.000135	-1.43E+15	-4.8E+14	-1.43E+15	-4.78E+14
b 2	Mismanaged plastic waste per capita person (kg per year)	2.41E+13	1.24E+13	1.95	0.054402	-4.64E+11	4.86E+13	-4.64E+11	4.86E+13
b 3	Mismanaged waste emitted to the ocean (kg)	172771.9	10501.8	16.5	2.02E-29	151920.36	193623.5	151920.4	193623.5
b 4	Amount of plastic emitted to ocean (kg)	-190.9364	14.03075	-13.6	6.04E-24	-218.7948	-163.078	-218.7948	-163.078
	p-value of beta 3 > 0.05, reject b 3								
	Remove (Mismanaged plastic waste per capita-person (kg per year))								

After removing the independent variables higher than 0.05, the R-squared value in the multiple regression statistics in Table IV drops a bit below the R-squared value of Table I. An explanation can be given by the R-squared value of 0.79, which indicates that the independent variables in this model can explain 79% of the variability in the dependent variable. Table V shows the results of the ANOVA table with a significant F value less than the significance level (0.05). This means that the overall model successfully rejects the null hypothesis and concludes that the model's results are statistically significant.

TABLE IV
MULTIPLE REGRESSION STATISTICS (MODEL 2)

Regression Statistics	
Multiple R	0.887913964
R Square	0.788391207
Adjusted R Square	0.781708824
Standard Error	8.66842E+14
Observations	99

TABLE V
ANOVA TABLE (MODEL 2)

ANOVA	**df**	**SS**	**MS**	**F**	**Significance F**	
Regression	3	2.65957E+32	8.86524E+31	117.980549	6.40657E-32	<0.05
Residual	95	7.13844E+31	7.51415E+29			
Total	98	3.37341E+32				

Table VI shows the regression output table of model 2 after removing the independent variables, which is higher than 0.05. The important thing is that all the p-values of independent variables are less than 0.05. Therefore, all coefficients are statistically at a 5% level of significance, so successful in rejecting the null hypothesis. According to this result, the best model in this multiple regression analysis can be obtained by summing the total value of coefficients.

TABLE VI
INTERPRET REGRESSION COEFFICIENTS TABLE (MODEL 2)

		Coefficients	**Standard Error**	**t Stat**	**P-value**	**Lower 95%**	**Upper 95%**	**Lower 95.0%**	**Upper 95.0%**
b0	Intercept	3E+14	9.952E+13	2.7173	0.0078212	7.3E+13	4.68E+14	7.2857E+13	4.68E+14
b1	Mismanaged plastic waste to ocean per capita person (kg per year)	-6E+14	1.772E+14	-3.5797	0.0005441	-9.9E+14	-2.82E+14	-9.859E+14	-2.82E+14
b2	Mismanaged waste emitted to the ocean (kg)	175150	10582.926	16.55	9.733E-30	154140	196159.4	154139.877	196159.4
b3	Amount of plastic emitted to ocean (kg)	-194.99	14.077998	-13.851	1.587E-24	-222.941	-167.0441	-222.94078	-167.0441
					all p-value<0.05				

After removing (Mismanaged plastic waste per capita person (kg per year))
H0:B1=B2=B3=0
H1:at least one b is not 0
Significance F is < 0.05, all p-value<0.05, so success to reject H0
Conclusion: All coefficients are statistically significant at a 5% significance level.

Another helpful outcome that can be obtained is the prediction macro in Figure 5. This prediction macro is obtained from the coefficients of intercept plus the coefficients of each independent variable multiple with a value based on the Interpret regression coefficients table. Furthermore, this multiple regression model predicts the total amount of MMP waste based on independent variables such as MMP waste in the ocean, mismanaged waste emitted into the ocean, and the amount of plastic emitted into the ocean. It can be used to forecast the future amounts of total MMP waste based on the trends of independent variables to plan or prepare the solution to slow down the growth of MMP waste.

Prediction Macro

MMP waste to ocean	Mismanaged waste emitted to ocean	Amount of plastic emitted to ocean	Total MMP waste
100	100	100	-6.31483E+16

Fig. 5 Prediction macro (model 2)

IV. CONCLUSION

Overall, this paper provides valuable insights into the consequences of ocean plastic pollution based on some analytical results. The variables used for analysis in this paper are humans can control and reduce so that humans can easily understand cognition. Three different types of dependent and independent variables of linear regression successfully display a strong relationship between each other. This means that the two variables of these three different types of linear regression will influence each other and reveal the impact of human mismanagement on plastic pollution into ocean. Therefore, humans must act the primary responsibility for ocean plastic pollution because all the mismanaged waste will cause by humans. Furthermore, multiple regression demonstrates a high R-Squared result on the independent variables collectively explaining the variation in the dependent variable. Moreover, a prediction macro was successfully created to predict the dependent variable (total mismanaged plastic waste) caused by independent variables.

As can be seen, this paper provides some factors related to plastic waste that cause ocean pollution by performing linear regression and multiple regression. This is to prove that the factors used are closely related to ocean pollution and that plastic pollution can be reduced as much as possible within people's control. All in all, the impact of these mismanaged plastic waste variables used in this paper is closely related to ocean pollution. More specifically, human behavior contributes to an annual increase in ocean plastic pollution. Therefore, humans must have better insights into cultivating sustainable solutions and maintaining ocean health to reduce or slowly pollute the ocean. In addition, the relevant world agencies must implement some actions to prevent ocean plastic pollution from continuing to increase in the future.

ACKNOWLEDGEMENT

This research was funded by Telekom Malaysia Research and Development, RDTC/221073 (MMUE/230002) and Multimedia University.

REFERENCES

[1] M. B. Tekman, "Impacts of plastic pollution in the oceans on marine species, biodiversity and ecosystems," *Impacts of Plastic Pollution in the Oceans on Marine Species*, Biodiversity and Ecosystems, Berlin, Germany: WWF Germany, 2022, pp. 1–221. doi:10.5281/zenodo.5898684.

[2] S. R. Seyyedi, E. Kowsari, S. Ramakrishna, M. Ghaebi, and A. Chinnappan, "Marine plastics, circular economy, and artificial intelligence: A comprehensive review of challenges, solutions, and policies," *Journal of Environmental Management*, vol. 345, p. 118591, Nov. 2023, doi: 10.1016/j.jenvman.2023.118591.

[3] C. J. Rhodes, "Plastic Pollution and Potential Solutions," *Science Progress*, vol. 101, no. 3, pp. 207–260, Sep. 2018, doi:10.3184/003685018x15294876706211.

[4] J. R. Jambeck et al., “Plastic waste inputs from land into the ocean,” *Science*, vol. 347, no. 6223, pp. 768–771, Feb. 2015, doi:10.1126/science.1260352.

[5] A. J. Mearns, A. M. Morrison, C. Arthur, N. Rutherford, M. Bissell, and M. A. Rempel‐Hester, “Effects of pollution on marine organisms,” *Water Environment Research*, vol. 92, no. 10, pp. 1510–1532, Oct. 2020, doi: 10.1002/wer.1400.

[6] C. Aguilera, A. Leija, M. Torres, and R. Mendoza, “Assessment of Environmental Quality in the Tamaulipas Laguna Madre, Gulf of Mexico, by Integrated Biomarker Response Using the Cross-Barred Venus Clam Chione elevata,” *Water, Air, & Soil Pollution*, vol. 230, no. 2, Jan. 2019, doi: 10.1007/s11270-019-4078-0.

[7] P. J. Landrigan et al., “Human Health and Ocean Pollution,” *Annals of Global Health*, vol. 86, no. 1, p. 151, Dec. 2020, doi:10.5334/aogh.2831.

[8] P. J. Landrigan et al., “The Minderoo-Monaco Commission on Plastics and Human Health,” *Annals of Global Health*, vol. 89, no. 1, 2023, doi:10.5334/aogh.4056.

[9] O. Pereao, B. Opeolu, and O. Fatoki, “Microplastics in aquatic environment: characterization, ecotoxicological effect, implications for ecosystems and developments in South Africa,” *Environmental Science and Pollution Research*, vol. 27, no. 18, pp. 22271–22291, Apr. 2020, doi: 10.1007/s11356-020-08688-2.

[10] D. W. Cole et al., “Aquaculture: Environmental, toxicological, and health issues,” *International Journal of Hygiene and Environmental Health*, vol. 212, no. 4, pp. 369–377, Jul. 2009, doi:10.1016/j.ijheh.2008.08.003.

[11] V. F. Arijeniwa et al., “Closing the loop: A framework for tackling single-use plastic waste in the food and beverage industry through circular economy- a review,” *Journal of Environmental Management*, vol. 359, p. 120816, May 2024, doi: 10.1016/j.jenvman.2024.120816.

[12] L. Zimmermann, S. Göttlich, J. Oehlmann, M. Wagner, and C. Völker, “What are the drivers of microplastic toxicity? Comparing the toxicity of plastic chemicals and particles to Daphnia magna,” *Environmental Pollution*, vol. 267, p. 115392, Dec. 2020, doi:10.1016/j.envpol.2020.115392.

[13] R. Shanker et al., “Plastic waste recycling: existing Indian scenario and future opportunities,” *International Journal of Environmental Science and Technology*, vol. 20, no. 5, pp. 5895–5912, Apr. 2022, doi:10.1007/s13762-022-04079-x.

[14] Ö. Bodin, “Collaborative environmental governance: Achieving collective action in social-ecological systems,” *Science*, vol. 357, no. 6352, Aug. 2017, doi: 10.1126/science.aan1114.

[15] J. S. Parra Sánchez, A. I. Oviedo Carrascal, and F. O. Amaya Fernández, “Analítica de datos: incidencia de la contaminación ambiental en la salud pública en Medellín (Colombia),” *Revista de Salud Pública*, vol. 22, no. 6, pp. 1–9, Nov. 2020, doi:10.15446/rsap.v22n6.78985.

[16] S. Gupta, D. Aga, A. Pruden, L. Zhang, and P. Vikesland, “Data Analytics for Environmental Science and Engineering Research,” *Environmental Science & Technology*, vol. 55, no. 16, pp. 10895–10907, Aug. 2021, doi: 10.1021/acs.est.1c01026.

[17] X. Liu, D. Lu, A. Zhang, Q. Liu, and G. Jiang, “Data-Driven Machine Learning in Environmental Pollution: Gains and Problems,” *Environmental Science & Technology*, vol. 56, no. 4, pp. 2124–2133, Jan. 2022, doi: 10.1021/acs.est.1c06157.

[18] Y. Zhu and Z. Chen, “Development of a DREAM-based inverse model for multi-point source identification in river pollution incidents: Model testing and uncertainty analysis,” *Journal of Environmental Management*, vol. 324, p. 116375, Dec. 2022, doi:10.1016/j.jenvman.2022.116375.

[19] Y. Zhu, H. Cao, Z. Gao, and Z. Chen, “A DiffeRential Evolution Adaptive Metropolis (DREAM)-based inverse model for continuous release source identification in river pollution incidents: Quantitative evaluation and sensitivity analysis,” *Environmental Pollution*, vol. 347, p. 123448, Apr. 2024, doi: 10.1016/j.envpol.2024.123448.

[20] Y. Zhu, Z. Chen, and Z. Asif, “Identification of point source emission in river pollution incidents based on Bayesian inference and genetic algorithm: Inverse modeling, sensitivity, and uncertainty analysis,” *Environmental Pollution*, vol. 285, p. 117497, Sep. 2021, doi:10.1016/j.envpol.2021.117497.

[21] Z. Steinmetz and H. Schröder, “Plastic debris in plastic-mulched soil—a screening study from western Germany,” *PeerJ*, vol. 10, p. e13781, Jul. 2022, doi: 10.7717/peerj.13781.

[22] D. Briassoulis, E. Babou, M. Hiskakis, and I. Kyrikou, “Analysis of long-term degradation behaviour of polyethylene mulching films with pro-oxidants under real cultivation and soil burial conditions,” *Environmental Science and Pollution Research*, vol. 22, no. 4, pp. 2584–2598, Sep. 2014, doi: 10.1007/s11356-014-3464-9.

[23] J. Dutta, “Plastic Pollution: A Global Problem from a Local Perspective,” *Open Access Journal of Waste Management & Xenobiotics*, vol. 1, no. 1, 2018, doi: 10.23880/oajwx-16000102.

[24] Q. Zhang et al., “Distribution and sedimentation of microplastics in Taihu Lake,” *Science of The Total Environment*, vol. 795, p. 148745, Nov. 2021, doi: 10.1016/j.scitotenv.2021.148745.

[25] Z. Zhang, Zulpiya·Mamat, and Y. Chen, “Current research and perspective of microplastics (MPs) in soils (dusts), rivers (lakes), and marine environments in China,” *Ecotoxicology and Environmental Safety*, vol. 202, p. 110976, Oct. 2020, doi:10.1016/j.ecoenv.2020.110976.

[26] V.-G. Le et al., “Review on personal protective equipment: Emerging concerns in micro(nano)plastic pollution and strategies for addressing environmental challenges,” *Environmental Research*, vol. 257, p. 119345, Sep. 2024, doi: 10.1016/j.envres.2024.119345.

[27] J. F. Provencher et al., “A Horizon Scan of research priorities to inform policies aimed at reducing the harm of plastic pollution to biota,” *Science of The Total Environment*, vol. 733, p. 139381, Sep. 2020, doi:10.1016/j.scitotenv.2020.139381.

[28] M. MacLeod, H. P. H. Arp, M. B. Tekman, and A. Jahnke, “The global threat from plastic pollution,” *Science*, vol. 373, no. 6550, pp. 61–65, Jul. 2021, doi: 10.1126/science.abg5433.

[29] B. Jiang, J. Yu, and Y. Liu, “The Environmental Impact of Plastic Waste,” *Journal of Environmental & Earth Sciences*, vol. 2, no. 2, pp. 26–35, Nov. 2020, doi: 10.30564/jees.v2i2.2340.