Results in Nonlinear Analysis 7 (2024) No. 1, 24–34
https://doi.org/10.31838/rna/2024.07.01.004
Available online at www.nonlinear-analysis.com

Results in Nonlinear Analysis
ISSN 2636-7556

Peer Reviewed Scientific Journal

Prediction the biodegradation rate of soil contaminated with different oil concentrations

Andrey Lipatov[1*], Elvira Belyanova[2], Irina Petunina[3]

[1]*Department of Management of Transport Systems, State University of Management, Moscow, Russia;* [2]*Department of Fundamental Sciences, Bauman Moscow State Technical University, Moscow, Russia;* [3]*Department of Higher Mathematics, Federal State Budgetary Educational Institution of Higher Education Kuban State Agrarian University named after I.T. Trubilin, Krasnodar, Russia*

Abstract

A multiple linear regression model is a practical statistical model for estimating relationships between a continuous dependent variable and predictor variables. The model itself is linear in that it consists of additive terms, each representing a predictor multiplied by an estimate of the coefficient. In addition, a constant (free term) is usually added to the model as well. Using two bacterial species (wild and recombinant), this paper develops a multiple model linear regression model to predict the biodegradation rate of soil contaminated with different oil concentrations. According to experimental findings, using a conventional microbial strain resulted in a 26% biodegradation of oil products after 6 weeks of incubation, whereas using the recombinant strain resulted in a 93% biodegradation. Factors such as crude oil concentration, number of days of incubation, and type of microbial strain were discovered to significantly influence the biodegradation rate based on visual and mathematical analysis. Mathematical models were developed to predict the biodegradation rate. The equation developed using multiple linear regression predicted the biodegradation rate with a coefficient of determination R2 = 0.549. The equation developed using polynomial regression predicted the biodegradation rate with a coefficient of determination R2 = 0.799. The resulting equations can be used to understand the relationship between the variables and also to predict the biodegradation rate of petroleum products.

Key words and phrases. biodegradation, oil, pollution, regression, soil, visual and mathematical analysis.

Email address: andrlipatov3@rambler.ru (Andrey Lipatov)

Received October 14, 2023; Accepted November 20, 2023; Online December 2, 2023

1. Introduction

Oil is currently one of the most commonly used raw materials for a wide variety of products required for human life [1]. However, the widespread use of oil and petroleum products is linked to spills, inevitably leading to soil contamination. Particularly, soil contamination is caused by oil well drilling, storage of oil products, and their processing, transportation, and sale [2,3].

In turn, soil contaminated by oil products is extremely hazardous to human health as well as plant and animal life [4,5]. Oil spills cause various environmental issues, including plant growth and development delays. Furthermore, oil spills cause insufficient soil aeration, reducing its mechanical strength and rendering it unsuitable for construction work. Large amounts of aliphatic and aromatic hydrocarbons and other organic compounds are present in oil and petroleum products [6,7]. Therefore, the US Environmental Protection Agency has classified these substances as a priority environmental pollutant, including a groundwater pollutant [8].

Bioremediation is a technique for cleaning oil and petroleum-contaminated soil using microorganisms. Bioremediation is a cost-effective, versatile, and efficient alternative to worldwide physical and chemical soil treatment [9,10].

There has been almost no research on mathematical modelling (including regression relationships) of bioremediation rates for soil contaminated with oil and oil products to date. It is important to keep in mind that regression models are empirical models or approximation functions that are utilised in circumstances in which there are several independent variables. Regression models are used when the true functional relationship between the dependent and independent variables is unknown. Zhu et al. [11] emphasises that regression models use complex forms of independent variables to represent the true unknown functions accurately.

1.1. Literature Review

Multiple Linear Regression (MLR) and Support Vector Machine (SVM) algorithms were used to predict the initial and final biodegradation rate based on a dataset of 171 chemicals. An MLR model was used in these works for a dataset containing molecules with fewer than nine carbon atoms. The authors hoped to discover linear relationships between chemical structure and biological reactions by constructing Quantitative Structure-Activity Relationships (QSAR) using MLR models [8,12]. Eventually, an empirical model was developed to represent polymer biodegradation, an exponential function with two different constants and a variable related to the biodegradable polymer's weight loss. The authors used the MATLAB Simulink platform to run the simulation. The developed mathematical models have been shown in works of Benyahia et al. [9] and Dragomir et al. [13] to simulate the experimental data satisfactorily.

The decomposition of crude oil in soil was studied using kinetic and statistical analyses in the works of Kheirkhah et al. [10] and Ani et al. [14]. The authors quantified the influence of various factors (including temperature) on the decomposition process using first- and second-degree kinetic models and single-factor analysis of variance.

A recent study Idroes et al. [15] compared linear and non-linear methods for predicting the Kovats retention index for 126 compounds extracted from the *Lippia origanoides* plant. Each compound's retention index was predicted using its molecular descriptors. There were 189 molecular descriptors for each compound in this study, and the best descriptors were chosen using a genetic algorithm (GA). It successfully selected the top five descriptors for use in constructing a multiple linear regression (MLR) model and an artificial neural network (ANN). Using MLR, the coefficient of determination R^2 was 0.959, 0.946, and 0.955 in training, validation, and testing, respectively, and the root-mean-square error (RMSE) was 48.00, 50.84, and 47.19. Meanwhile, ANN yielded R^2 values of 0.963, 0.947, and 0.962 and RMSE values of 45.45, 50.59, and 43.20. Compared to MLR, the ANN model performed better, with R2 increasing by 0.004, 0.001, and 0.007 and RMSE decreasing by 2.55, 0.25, and 3.99. Based on the obtained predictive results, it is known that the ANN and MLR methods produce nearly comparable results.

Using optimal molecular descriptor data, Idroes et al. [16] determined Kovats retention indices for 51 flavour and aroma compounds. The optimal molecular descriptor was chosen using a genetic algorithm written in Perl. The optimal molecular descriptor was used to predict the Kovats retention index using multiple linear regression generated with R. It was demonstrated that the molecular descriptor value could be determined efficiently using the free (open source) software Online Chemical Database. According to the findings, 51 flavour and aroma compounds yield 170 molecular descriptors. Six optimal molecular descriptors were chosen from these molecular descriptors based on 200 repetitions to construct the MLR. The best model had the following attributes: optimisation parameter R^2 =0.981, corrected value R^2=0.978, and RMSE=43.50.

Benkachcha et al. [17] has demonstrated how crucial predicting is to supply chain management. It enables businesses to foresee their customers' needs and adapt accordingly. Research findings on how ANN and MLR can enhance the accuracy of predictions are presented in the article [17]. The findings indicate that ANN- and MLR-based models are highly promising for predicting problems.

Problem Statement. Based on the literature review performed, it can be concluded that the forecast of the level of soil pollution with oil products is very relevant. In addition, the choice of factors that have the greatest impact on the degree of pollution, as well as the effectiveness of using linear or polynomial regression to predict the level of pollution, remains open.

Based on multiple linear and polynomial regression models, this study aims to identify significant variables in the efficiency of the biodegradation process of oil-contaminated soil.

To this end, the following objectives must be met:

1. Selection of factors affecting the biodegradation process of oil and petroleum-contaminated soil.
2. Mathematical and visual analysis of experimental data on biodegradation of oil and petroleum-contaminated soil.
3. Performance comparison of the developed multiple linear correlation and polynomial regression models.

This article is further structured based on the purpose and objectives as follows: Section 2 – Methods and Materials – discusses multiple linear regression and polynomial regression theoretically; Section 3 – Results – presents the findings of experimental studies as well as developed models of multiple linear regression and polynomial regression; Section 4 – Discussion – covers the findings obtained and compares them to those obtained by other authors, and Section 5 – Conclusions – discusses the key conclusions and proposals based on the research findings.

2. Materials and Methods

The authors used data from the study of Ajona and Vasanthi [18], which presents the findings of studies to assess the effect of local microorganisms on hydrocarbon biodegradation, with microbial strains isolated from a crude oil-contaminated region. A recombinant process was performed on selected local microorganisms during the procedure. These wild and recombinant microbial strains were inoculated in soil with various oil concentrations (0.5%, 1%, 3%, 5%, and 7%) for 42 days, and the total petroleum hydrocarbon content was determined. The biodegradation process was studied after 42 days of incubation, and the results revealed that the recombinant microbial strain outperformed the wild microbial strain in terms of biodegradation. In the abovementioned study, forty samples with varying oil concentrations and incubation times (0, 14, 28, and 42 days) were obtained for both native and recombinant microbial strains. The biodegradation rate for each sample was recorded [11,18].

Indicators TPH and BD were determined as follows. 10 g of artificially polluted soil samples were collected at predetermined intervals of (1) zero weeks, (2) two weeks, (3) four weeks, and (4) six weeks TPH, GC-MS analysis for biodegradation monitoring. For the intentionally contaminated soil samples, hydrocarbon (TPH) calculations were performed as follows. 10 g of each batch of artificially polluted soil sample was dissolved in 50 ml hexane and agitated for ten minutes using a mechanical shaker. The solution was filtered from each sample using a 1 μl Whatman filter paper for GC-MS. The extracted oil samples were GC-MS analyzed for hydrocarbon biodegradation levels (TPH). By comparing the zero-week sample areas of the peaks taken from treated samples, the percentage loss of biodegraded alkanes was calculated. The total concentration of hydrocarbons in the extracted crude oil from the microcosms before and after treatment with the two bacterial isolates, wild and recombinant, was monitored and recorded using GC-MS at each specified time interval. The percentage of deterioration was estimated using the following expression:

$$BD = \left[\frac{TPH_{initial} - TPH_{treated}}{TPH_{initial}}\right] \times 100 \tag{1}$$

Regression analysis is a quantitative research technique primarily used to examine the relationship between dependent and independent variables. The nature of the relationship between independent parameters like crude oil concentration, number of incubation days, and microbial strain type and the dependent parameter, biodegradation rate, was investigated in this study [11,18].

Each independent variable was first visualised in relation to the dependent variable to determine how they interacted. Then, the data were visualised to make it easier to spot patterns and random points, ultimately leading to the discovery of additional information. Furthermore, the data sets should be checked for linearity, absence of collinearity, homoscedasticity, and the presence of a normal distribution [19].

MLR and polynomial regression (PR) were the two regression models used in this study to determine the relationship between the dependent and independent variables. The regression algorithm's purpose was to plot the optimal fit line or curve between the data [20]. The entire dataset was divided into a training portion comprising 80% and a test portion comprising 20%. The regression equation was built with the training set and tested with the test set. A flowchart describing this study is shown in Figure 1.

2.1. Multiple Linear Regression

Multiple linear regression is an extension of linear regression (also known as least squares). MLR employs multiple independent variables to predict the outcome of the dependent variable [21,22]. The underlying functional relationship between the dependent and independent variables is unknown, but using complex forms of the independent variables, the MLR model accurately approximates the

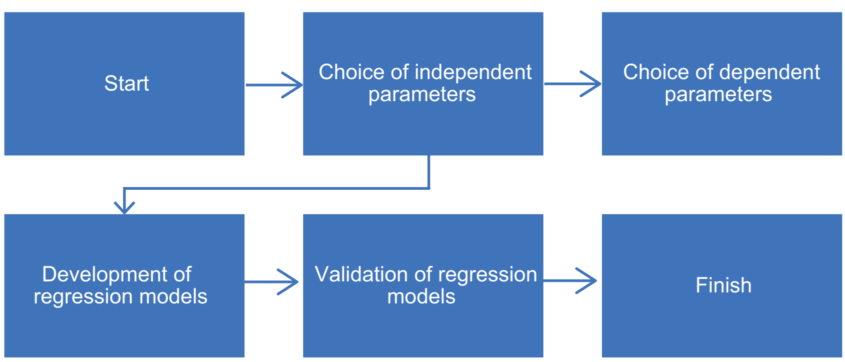

Figure 1. Mathematical Model Development Flowchart (developed by the authors).

true unknown function [7,23]. A multiple linear regression model for n independent variables (X) with unknown parameters (β) and dependent variables (Y) is expressed as follows:

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots \beta_n x_n \tag{2}$$

Where $\hat{y}$ is the predictive response variable that changes as the variables x_1, x_2,...x_n change. The parameter β_0 is where the regression line meets the dependent variable when all independent variables equal zero. The parameters β_1, β_2,...β_n are the estimated regression coefficients that quantify the relationship between the independent variables x_1, x_2,...x_n, and the result, respectively.

2.2. Polynomial Regression

Python programming describes the relationship between an independent variable x and dependent variable y as an nth-degree polynomial by x for polynomial regression. $E(y\,|\,x)$ represents the non-linear relationship between the value and the conditional mean . Although PR fits a non-linear model to the data, this is a linear statistical estimation problem since the regression function is linear with respect to the unknown parameters inferred from the data. Therefore, PR is considered a subset of multiple linear regression [13], [14], [24], [25]. The general form of a nth-degree polynomial regression equation with one independent variable is as follows:

$$\hat{y} = \beta_0 + \beta_1 x + \beta_2 x^2 + \ldots \beta_n x^n \tag{3}$$

Where $\hat{y}$ is the expected response variable, and x is the independent variable.

In some cases, such as this one involving the MLR model, the optimal line (i.e. a first-degree polynomial) is insufficient. Typically, the second- (or higher) degree polynomial can fit curves to calibration data.

3. Results

Figure 2 illustrates the relationship between crude oil concentration and biodegradation rate. The graph indicates that as concentration levels rise, biodegradation significantly decreases. Moreover, linearly, this means a downward trend. With an increase in oil concentration from 0.5 to 7%, the degree of biodegradation rate from ~95% to ~20%.

The effect of the number of incubation days on the biodegradation rate is shown in Figure 3. Clearly, there is an upward trend here. The biodegradation rate increased sharply from 0 to ~95 % with the number of days (from 0 to 42).

Wild and recombinant microbial strains are the two organisms under study of Zhu et al. [11], Ajona and Vasanthi [18]. Because these microbial strains are categorical data, they have been assigned numerical values 1 and 2 to represent wild and recombinant microbial strains, respectively. Figure 4 shows how the biodegradation rate changes depending on the microbial strain (organism). Microbial strain no. 2 has much more biodegradable rate (~95 %) in comparison with microbial strain no. 1 (~25 %).

Table 1 shows the wild and recombinant organisms' total petroleum hydrocarbon (TPH) and biodegradation potential (BD) content.

The data shown in Figure 4 and Table 1 show that the recombinant type has a higher biodegradation rate than the wild microbial strain.

The recombinant microbial strain is discovered to perform better than the wild strain in terms of biodegradation rate. Furthermore, the biodegradation rate increases with lower concentrations and longer incubation times.

The Python-programmed MLR model used in this study is calculated from a training dataset and has the following structure:

$$BD = -21.699 - 4.423C + 0.928D + 28.628M \tag{4}$$

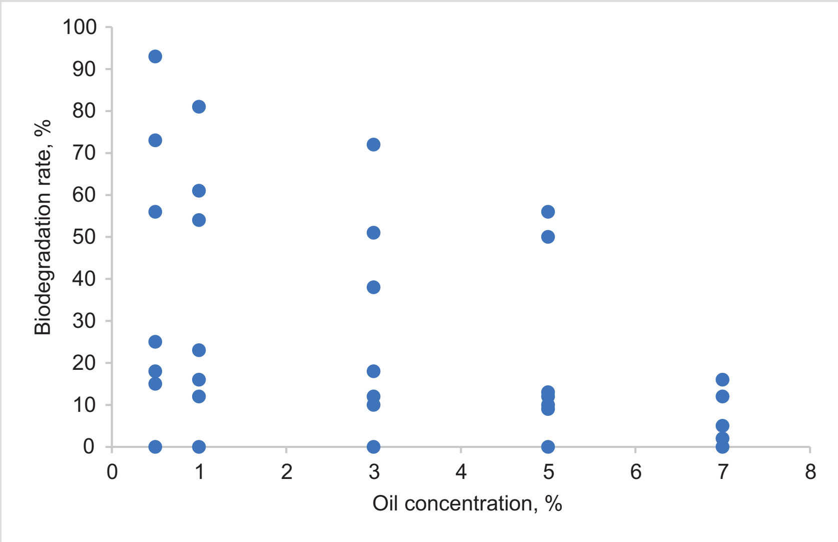

Figure 2. Biodegradation Rate vs Oil Concentration (developed by the authors based on [17]).

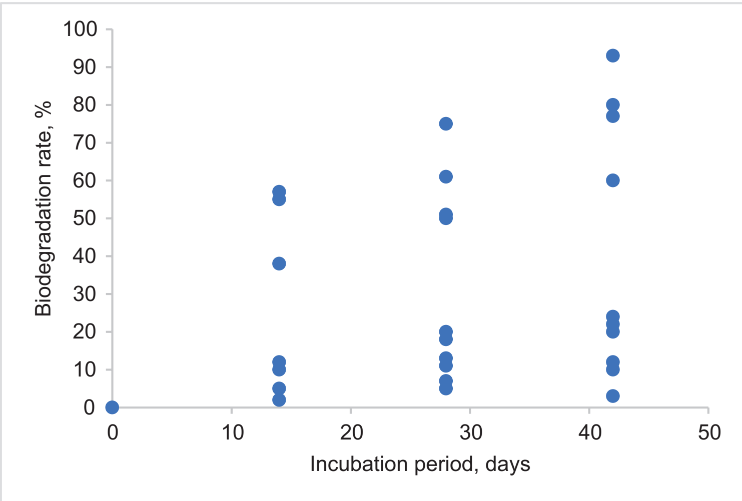

Figure 3. Biodegradation Rate vs Incubation Period (developed by the authors based on [17]).

Where BD is the expected biodegradation rate, C is the crude oil concentration, D is the number of incubation days, and M is the microbial strain type, which takes values 1 and 2 to represent wild and recombinant microbial strains. The coefficient of determination of the MLR model (R^2) is 0.549. The low R^2 value explains why highly variable data points are scattered around the regression line. Hence, relying on the resulting MLR equation for accurate predictions is impossible.

A training dataset is used to examine the relationship between the dependent and independent variables (80% of the total data) to develop a polynomial regression model. As a result, the following PR equation is developed:

$$BD = 24.324 + 14.827C + 3.423D + 5.739M - 1.659C^2 - 9.312CD - 20.745CM - 7.672D^2 + 27.562DM + 5.858M^2 \tag{5}$$

Where BD is the expected biodegradation rate, C is the crude oil concentration, D is the number of incubation days, and M is the microbial strain type, which takes values 1 and 2 to represent wild and recombinant microbial strains. The PR model's R^2 value is 0.799, indicating that most of the

Figure 4. Biodegradation Rate vs Microorganism Type (developed by the authors based on [17]).

Table 1. TPH and BD of Wild and Recombinant Strains.

Time, weeks	Wild organism		Recombinant organism	
	TPH	BD, %	TPH	BD, %
0	3.91E+08	-	3.8E+08	-
2	3.35E+08	13	1.8E+08	57
4	3.17E+08	20	1.2E+08	72
6	2.88E+08	26	3.2E+08	93

Source: Developed by the authors based on [18].

variation in the response variable is concentrated at its mean value. Generally, the higher the R^2, the better the regression model fits the data.

4. Discussion

According to the findings, laboratory soil samples containing various microbial strains were used to determine TPH biodegradation. TPH biodegradation was especially rapid during the first two to four weeks of incubation. The pattern of crude oil decomposition in the soil revealed a rapid early decomposition stage followed by a period of slight concentration change. Samples were collected and analysed every fourteen days. Both bacteria's biodegradation capacity was assessed. The recombinant bacteria performed better than the wild bacteria. Two regression models were developed to simulate the biodegradation rate of crude oil using wild and recombinant microbial strains.

To compare our results with the results obtained by other authors, we can use the data presented in the study of Michael-Igolima et al. [26]. In this work, it was shown that biological methods for cleaning soil from oil pollution are quite effective and have the prospect of further practical application. Particularly, Rodriguez-Campos et al. [27] conducted bioremediation study of hydrocarbon contaminated soil using vermremediation, phytoremediation, and bioaugmentation individually and in combinations. Results obtained after 112 days showed that the 3 techniques were able to degrade hydrocarbon in the soil. However, the highest TPH removal was with earthworms and bacteria (86.4%), followed by earthworms plus plants plus bacteria (82.7%), and bacteria (82.6%). Biological method can be used in natural attenuation after chemical or physical remediation to restore soil flora and fauna. Valderrama et al. [28] reported 80% degradation of PAH (polycyclic aromatic hydrocarbons) in soil by chemical oxidation, however, the maximum cleanup efficiency was achieved by combining chemical and biological methods. Similarly, Tsai and Kao [29] reported 90% clean-up efficiency by combined methods of chemical and biological remediation. The studies show that combined remediation is an effective soil remediation method, integrating biological methods in the treatment train enhances the remediation process and restores soil flora and fauna.

Before employing any statistical analysis, it is critical to understand its origin, scope, and limitations [18,30]. The assumptions that must be considered when employing linear regressions were examined and supported by data. It is noteworthy that despite polynomial regression accommodating a non-linear model with the data, the regression function is linear with respect to the unknown parameter obtained from the data. Consequently, it is deemed linear in the context of statistical estimation. Data visualisation aids in testing the assumptions accounted for in linear regression.

As in this case, multiple linear regression models are frequently used to predict certain crude oil and petroleum product properties.

A multiple linear regression method was used in the study of Hussein and Abdula [31] to predict vitrinite reflectance from acoustic and resistivity logs in the Shaykhan-2 well for the Sargelu and Naokelekan formations. The oil matrix rock that reached the oil window and produced hydrocarbons was anticipated to have higher acoustic and resistivity log readings than the oil matrix rock that is still in the diagenesis stage and has not produced oil or gas. The presence of a conductive oil phase (pyrobitumen) and an increase in residual water salinity due to the solubility of water vapour in the produced gas are two possible explanations for the decrease in resistivity with increasing maturity. A multiple linear correlation model was used to achieve the best agreement between the vitrinite reflectance measured on cuttings samples and the vitrinite reflectance predicted from logging data in the Shaykhan-2 well.

The concentration of oxidised fatty acids (OFA) in first-press olive oil samples with varying oxidative status was estimated using Fourier-transform infrared spectroscopy (FTIR) followed by multivariate spectral data processing [32]. The oils' FTIR spectra (4000–700 cm-1) were divided into 25 wavelength regions. The normalised absorbance of the peak areas in these regions was used as a predictor. MLR models were developed to predict OFA concentration. MLR model with eight predictors could predict OFA concentration with an average error of 17% after the data were transformed from the cubic root. Since they are more prone to oxidation, the main wavelength regions chosen to construct this MLR model corresponded to the C–H double bond (trans and cis, stretch), –C–H (CH2, stretch asym), O–H (in-plane bending), C–O (stretch), –H–C double bond C–H– (cis?) and double bond CH2 (wagging).

The study of Ozturk and Basar [33] aims to reduce ship-generated air pollution and shipping operating costs by implementing measures to improve voyage management efficiency. This study's methodological approach is based on decision support systems (DSSs). DSSs were created using Multiple Linear Regression Analysis (MLRA) forecasting methods and ANN for Fuel Oil Consumption (FOC). FOC forecast models were developed using voyage report data from 19 container vessels, including revolutions per minute (RPM), pitch, average draught, trim, weather conditions, and FOC variables. The FOC forecasting models' compatibility values (76–90%) were satisfactory. The developed models compared the performance of MLRA and ANN methods for FOC prediction and showed the effect of RPM, trim, ballast, and weather conditions optimisation methods on energy efficiency. The results show that optimising RPM, trim, weather routing, and ballast can save 32–37%, 6.5–8%, 7–12%, and 6–8%, respectively.

Regression equations not only aid in predicting the response variable but also in comprehending the relationship between the independent variables and how they influence the dependent variable. According to the regression models and graphs, the recombinant microbial strain is more biodegradable than the wild strain. Additionally, the biodegradation rate is negatively impacted by crude oil concentration while positively impacted by incubation days. Finally, the recombinant microbial strain with the least amount of crude oil and the longest incubation time demonstrated maximum biodegradation.

The following are some observations and suggestions for future research based on this study:

1. Understanding the structure and natural balance of existing bacteria in contaminated areas is fundamental to developing new bioremediation techniques as a functional tool for cleaning up crude oil-contaminated environments.
2. Identifying and using oil-feeding bacteria will aid in bioremediation techniques. This cost-effective and environmentally friendly method can be implemented at all oil-contaminated sites, thereby helping to protect the environment.

5. Conclusions

This study identified significant variables for the efficiency of the bioremediation process of oil-contaminated soil based on multiple linear and polynomial regression models.

To accomplish this, experimental data were analysed using microbial strains from an area contaminated with crude oil to determine how local microorganisms influence hydrocarbon biodegradation. These wild and recombinant microbial strains were exposed to various oil concentrations (0.5%, 1%, 3%, 5%, and 7%) in soil for 42 days, and the total petroleum hydrocarbon content was determined. The biodegradation process was studied after 42 days of incubation. The study described above yielded 40 samples with varying oil concentrations and incubation times (0, 14, 28, and 42 days) for both the local and recombinant microbial strains. The corresponding biodegradation rate was recorded for each sample.

It was determined through mathematical and visual analysis that variables like crude oil concentration, the number of incubation days, and the type of microbial strain significantly impacted the biodegradation rate.

According to experimental findings, using a conventional microbial strain resulted in a 26% biodegradation of oil products after 6 weeks of incubation, whereas using the recombinant strain resulted in a 93% biodegradation.

Mathematical models were developed to predict the biodegradation rate. Using multiple linear regression, an equation was made that could predict the biodegradation rate with a determination coefficient of R2 = 0.549. Using polynomial regression, the same equation could predict the biodegradation rate with a determination coefficient of R2 = 0.799.

The resulting equations can be used to understand the relationship between the variables and predict the biodegradation rate.

Acknowledgments

Not applicable.

Conflict of interest

The authors declare there is no conflict of interest.

References

[1] Y. Nan, R. Sun, Z. Zhen, C. Fangjing, Measurement of international crude oil price cyclical fluctuations and correlations and correlation with the world economic cyclical changes, *Energy*, 2022, 260, 124946.

[2] E. Hauptfeld, J. Pelkmans, T.T. Huisman, A. Anocic, B.L. Snoek, F.A.B. von Meijenfeldt, J. Gerritse, J. van Leeuwen, G. Leurink, A. van Lit, R. van Uffelen, M.C. Koster, B.E. Dutilh, A metagenomic portrait of the microbial community responsible for two decades of bioremediation of poly-contaminated groundwater, *Water Res.*, 2022, 221, 118767.

[3] A. Koolivand, R. Saeedi, F. Coulon, V. Kumar, J. Villasenor, F. Asghari, F. Hesampoor, Bioremediation of petroleum hydrocarbons by vermicomposting process bioaugmented with indigenous bacterial consortium isolated from petroleum oily sludge, *Ecotoxicol. Environ. Saf.*, 2020, 198, 110645.

[4] R. Margesin, F. Schinner, Biodegradation and bioremediation of hydrocarbons in extreme environments, *Appl. Microbiol. Biotechnol.*, 2001, 56, 650–663.

[5] M.D. Yuniati, Bioremediation of petroleum-contaminated soil: A review, *IOP Conf. Ser.*: Earth Environ. Sci., 2018, 118, 1–7.

[6] Y. Lv, J. Bao, Y. Dang, D. Liu, T. Li, S. Li, Y. Yu, L. Zhu, Biochar aerogel enhanced remediation performance for heavy oil-contaminated doil via biostimulation strategy, *J. Hazard. Mater.*, 2023, 443, 130209.

[7] Z. Ines, B. Amina, R. Mahmoud, S.M. Dalila, Aliphatic and aromatic biomarkers for petroleum hydrocarbon monitoring in Khniss Tunisian Coast, (Mediterranean Sea), *Procedia Environ. Sci.*, 2013, 18, 211–220.

[8] USEPA, EPA Test Methods for Evaluating Solid Wastes. OSW Web Site Manager, 2001.

[9] F. Benyahia, M. Abdulkarim, A. Zekri, O. Chaalal, H. Hasanain, Bioremediation of crude oil contaminated soils a black art or an engineering challenge?, *Process Saf. Environ. Prot.*, 2005, 83, 364–370.

[10] T. Kheirkhah, P. Hejazi, A. Rahimi, Effects of utilising sawdust on non-ligninolytic degradation of high concentration of n-hexadecane by white-rot fungi: kinetic analysis of solid-phase bioremediation, *Environ. Technol. Innovat*, 2020, 19, 100887.

[11] L. Zhu, J.P. O'Dwyer, V.S. Chang, C.B. Granda, M.T. Holtzapple, Multiple linear regression model for predicting biomass digestibility from structural features, *Bioresour. Technol*, 2010, 101, 4971–4979.

[12] W. Tang, Y. Li, Y. Yu, Z. Wang, T. Xu, J. Chen, J. Lin, X. Li, Development of models predicting biodegradation rate rating with multiple linear regression and support vector machine algorithms, *Chemosphere* 2020, 253, 110645.

[13] T. Dragomir, A. Pana, V. Ordodi, V. Gherman, G. Dumitrel, S. Nanu, An empirical model for predicting biodegradation profiles of glycopolymers, *Polymers* 2021, 13, 1–13.

[14] K.A. Ani, E.C. Chukwuma, Kinetics and statistical analysis of the bio-stimulating effects of goat litter in crude oil biodegradation process, *Beni-Suef Univ. J. Basic Appl. Sci.,* 2020, 9, 29.

[15] R. Idroes, T.R. Noviandy, A. Maulana, R. Suhendra, N.R. Sasmita, M. Muslem, G.M. Idroes, P. Kemala, I. Irvanizam, Application of genetic algorithm-multiple linear regression and artificial neural network determinations for prediction of Kovats Retention Index, *Int. Rev. Model. Simul.*, 2021, 14, 137–145.

[16] R. Idroes, T.R. Noviandy, A. Maulana, R. Suhendra, N.R. Sasmita, M. Muslem, G.M. Idroes, I. Irvanizam, Retention index prediction of flavour and fragrance by multiple linear regression and the genetic algorithm, *Int. Rev. Model. Simul.*, 2019, 12, 373–380.

[17] S. Benkachcha, J. Benhra, H. El Hicham, Demand forecasting in supply Chain: Comparing multiple linear regression and artificial neural networks approaches, *Int. Rev. Model. Simul.*, 2014, 7, 279–286.

[18] M. Ajona, P. Vasanthi, Bio-remediation of crude oil contaminated soil using recombinant native microbial strain, *Environ. Technol. Innov.*, 2021, 23, 101635.

[19] S. Sharma, R. Gupta, R. Bhatia, A.P. Toor, H. Setia, Predicting microbial response to anthropogenic environmental disturbances using artificial neural network and multiple linear regression, *Int. J. Cogn. Comput. Eng.*, 2021, 2, 65–70.

[20] Y. Wang, S. Wu, H. Wang, Y. Dong, X. Li, S. Wang, H. Fan, X. Zhuang, Optimisation of conditions for a surfactant-producing strain and application to petroleum hydrocarbon-contaminated soil bioremediation, *Colloids Surf. B: Biointerfaces*, 2022, 213, 112428.

[21] X. An, B. Zhong, G. Chen, W. An, X. Xia, H. Li, F. Lai, Q. Zhang, Evaluation of bioremediation and detoxification potentiality for papermaking black liquor by a new isolated thermophilic and alkali-tolerant Serratia sp. *AXJ-M, J. Hazard. Mater.*, 2021, 406, 124285.

[22] E.E. Raimondo, J.D. Aparicio, A.L. Bigliardo, M.S. Fuentes, C.S. Benimeli, Enhanced bioremediation of lindane-contaminated soils through microbial bioaugmentation assisted by biostimulation with sugarcane filter cake, *Ecotoxicol. Environ. Saf.*, 2020, 190, 110143.

[23] Y.W. Chang, C.J. Hsieh, K.W. Chang, M. Ringgaard, C.J. Lin, Training and testing low-degree polynomial data mappings via linear SVM, *J. Mach. Learn. Res.*, 2010, 11, 1471–1490.

[24] V.A. Profillidis, G.N. *Botzoris, Modeling of Transport Demand*, Elsevier, 2018.

[25] G. Jiang, A. Thummala, M.V.S. Wadhwa, Applications of statistical regression and modelling in fill-finish process development of structurally related proteins, *J. Pharm. Sci.*, 2011, 100, 464–481.

[26] U. Michael-Igolima, S.J. Abbey, A.O. Ifelebuegu, A systematic review on the effectiveness of remediation methods for oil contaminated soils, *Environ. Adv.*, 2022, 9, 100319.

[27] J. Rodriguez-Campos, A. Perales-Garcia, J. Hernandez-Carballo, F. Martinez-Rabelo, B. Hernandez-Castellanos, I. Barois, S.M. Contreras-Ramos, Bioremediation of soil contaminated by hydrocarbons with the combination of three technologies: bioaugmentation, phytoremediation, and vermremediation, *J. Soils Sediments.*, 2019, 9, 1981–1984.

[28] C. Valderrama, R. Alessandri, T. Aumola, J.L. Cortina, X. Gamisans, T. Tuhkanen, Oxidation by Fenton`s reagent combined with biological treatment applied to a creosote-contaminated soil, *J. Hazard. Mater.,* 2009, 166 (2–3), 594–602.

[29] T.T. Tsai, C.M. Kao, Treatment of petroleum-hydrocarbon contaminated soils using hydrogen peroxide oxidation catalysed by waste basic oxygen furnace slag, *J. Hazard. Mater.,* 2009, 170 (1), 466–472.

[30] O.B. Said, C. Cravo-Laureau, F. Armougom, S. Cipullo, M.B. Khelil, M.B.H. Yahiya, A. Douihech, H. Beyrem, F. Coulon, R. Duran, Enhanced pilot bioremediation of oily sludge from petroleum refinery disposal under hot-summer Mediterranean climate, *Environ. Technol. Innov.*, 2021, 24, 102037.

[31] H.S. Hussein, R.A. Abdula, Multiple linear regression approach for the vitrinite reflectance estimation from well logs: A case study in Sargelu and Naokelekan Formations – Shaikhan-2 Well, Shaikhan oil field, Iraq, *Egypt. J. Pet.*, 2018, 27, 1095–1102.

[32] M.J. Lerma-Garcia, E.F. Simo-Alfonso, A. Bendini, L. Cerretani, Rapid evaluation of oxidised fatty acid concentration in virgin olive oil using Fourier-transform infrared spectroscopy and multiple linear regression, *Food Chem.,* 2011, 124, 679–684.

[33] O.B. Ozturk, E. Basar, Multiple linear regression analysis and artificial neural networks based decision support system for energy efficiency in shipping, *Ocean Eng.*, 2022, 243, 110209.