Machine Learning Predictions of Oil Yields Obtained by Plastic Pyrolysis and Application to Thermodynamic Analysis

Elizabeth R. Belden, Matthew Rando, Owen G. Ferrara, Eric T. Himebaugh, Christopher A. Skangos, Nikolaos K. Kazantzis, Randy C. Paffenroth, and Michael T. Timko*

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Chemical recycling via thermal processes such as pyrolysis is a potentially viable way to convert mixed streams of waste plastics into usable fuels and chemicals. Unfortunately, experimentally measuring product yields for real waste streams can be time- and cost-prohibitive, and the yields are very sensitive to feed composition, especially for certain types of plastics like poly(ethylene terephthalate) (PET) and polyvinyl chloride (PVC). Models capable of predicting yields and conversion from feed composition and reaction conditions have potential as tools to prioritize resources to the most promising plastic streams and to evaluate potential preseparation strategies to improve yields. In this study, a data set consisting of 325 data points for pyrolysis of plastic feeds was collected from the open literature. The data set was divided into training and test sub data sets; the training data were used to optimize the seven different machine learning regression methods, and the testing data were used to evaluate the accuracy of the resulting models. Of the seven types of models, eXtreme Gradient Boosting (XGBoost) predicted the oil yield of the test set with the highest accuracy, corresponding to a mean absolute error (MAE) value of 9.1%. The optimized XGBoost model was then used to predict the oil yields from real waste compositions found in Municipal Recycling Facilities (MRFs) and the Rhine River. The dependence of oil yields on composition was evaluated, and strategies for removing PET and PVC were assessed as examples of how to use the model. Thermodynamic analysis of a pyrolysis system capable of achieving oil yields predicted using the machine-learned model showed that pyrolysis of Rhine River plastics should be net exergy producing under most reasonable conditions.

KEYWORDS: pyrolysis, waste plastics, machine learning, regression analysis, chemical recycling
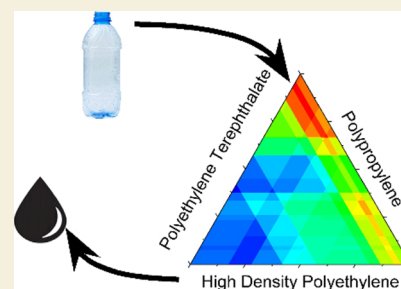
1. INTRODUCTION

Every year approximately 370 million tons of plastic are generated globally.[1] In the USA, approximately 9% of plastic is recycled, 16% is combusted with energy recovery, and the remaining 75% is landfilled.[2] In other locations, environmental disposal is common practice and combined with losses during transport and loss from landfills, results in an estimated release of 10−20 million tons of plastic per year into the world's oceans.[3] Once in the environment, waste plastics have negative impacts on plant and animal life[4] with eventual negative impacts on human health.[5]

Increased recycling is an obvious solution to the problem of environmental release. Although recent recycling initiatives have increased the amounts of plastic that are recycled, global recycling capabilities are limited by the need for single plastic streams for effective recycling and further sensitivity to the presence of contaminants like dyes, additives, and residual products.[6] Switching to biodegradable plastics has the potential to reduce the negative impacts of plastics released to the environment, but biodegradable plastics that duplicate the properties of synthetic ones are not available.[7] Without increasing the world's recycling capabilities or replacing synthetic plastics with biodegradable ones, landfilled wastes will continue to end up in the environment outside of landfills and in oceans and rivers globally.[8]

Reducing the amount of plastics that ends up in landfills and the environment is therefore a priority, and one area of specific concern is plastic that is transferred into rivers as it not only pollutes the rivers but is also transported to the world's oceans where it accumulates and becomes an environmental threat.[9,10] Recently, a handful of highly industrialized river systems, including the Yangtze, Ganges, and Xi, have been identified as especially problematic.[11] Because of the limitations of recycling and slow progress in replacing synthetic plastics with biodegradable versions, new and innovative technologies are needed to valorize waste plastics and reduce their flow into the environment.[12,13]

Different plastic valorization techniques have been developed over the years, falling mainly into two categories

mechanical and chemical recycling. A key advantage of chemical recycling techniques is they are more compatible with mixed waste plastics than mechanical recycling. The end products of chemical recycling are fuels, chemicals, or monomers, depending on the feed and the specific process technology.[14] Although many dozens of plastics are in use, many of these are used in low-volume, niche applications. Six plastic types account for >85% of all plastic used today,[15] meaning that reducing plastic waste should prioritize these plastics, which include poly(ethylene terephthalate) (PET), high-density polyethylene (HDPE), polyvinyl chloride (PVC), low-density polyethylene (LDPE), polypropylene (PP), and polystyrene (PS). Collectively, these plastics can be termed as the "big six" since they are denoted numerically in the U.S. recycling system in this order from one to six.[16]

Recent focus has shifted to circular economy for plastics with emphasis on "upcycling," a term used to describe processes that yield products that are more valuable than the virgin plastics themselves.[17] Chemical recycling can qualify as upcycling, in certain cases, especially when performed in the presence of a catalyst.[18] On the other hand, recycling to products with value equal to the virgin plastic or even down cycling to produce products of less value than the original plastic can play an important role. Products such as chemicals, fuels, and monomers are undoubtedly more valuable than waste plastics,[19] and this type of chemical recycling can help reduce plastics entering landfills or the environment. Unlike upcycled products, the market size for chemicals, fuels, and monomers is commensurate with the size of the waste plastic feed.

Of the different forms of chemical recycling, thermal depolymerization techniques, such as pyrolysis and hydro-thermal liquefaction (HTL), have been especially promising for their ability to handle mixed wastes, their compatibility with contaminates, and their ability to achieve high oil product yields without using a catalyst.[20] Pyrolysis heats plastics to high temperatures in the absence of oxygen to break the carbon–carbon bonds within polymers to return them to their monomer states.[21] HTL operates similarly to pyrolysis but in the presence of sub- or supercritical water.[22] Both pyrolysis and HTL have the ability to convert plastics into oils with conversion rates of greater than 90%, depending on the process conditions and the plastic feed.[23,24] Results like these have spurred significant research into thermal depolymerization of plastics.[14]

Although previous studies show the great promise of thermal depolymerization, a technological problem of pyrolysis reactor design is that different polymers have very different thermal reactivities. Accordingly, the composition of a given plastic waste stream significantly affects both the oil yield and the overall conversion.[25] Given the number of potential feed streams, an inability to predict oil yields hampers prioritization of resources to the situations that are most promising for investment. Furthermore, for every new feed of polymers, the operating parameter matrix must be reoptimized, which is time consuming and potentially cost-prohibitive, especially when the oil yields are unknown. An efficient way to determine the oil yield obtainable from a given plastic waste stream would permit allocation of finite resources to streams most likely to be thermodynamically and economically favorable.

Models that describe plastic depolymerization have been available for over 20 years.[26−28] These models are full kinetic networks that track polymer chain length using the method of moments and formation of key products using systems of ordinary differential equations.[26−28] While providing accuracy and reliability, full kinetic models are computationally expensive, require years to develop for new plastics or plastic mixtures, and are not available for anything more than binary plastic mixtures.[26−28] Fortunately, the level of resolution offered by full kinetic models is not always required for process design, opening an opportunity for lower-resolution yet accurate methods of predicting key reaction outcomes, such as pyrolysis oil yields.

Machine learning techniques have the potential to predict reaction results, such as yield and conversion, with significantly less development time and computational expense than full kinetic models. Unlike full kinetic models, which require detailed measurements of polymer chain length and key product and intermediate formation rates as a function of conversion, machine learning models use abundant historical data to train models capable of predicting outcomes of new situations.[29] Machine-learned models are black box, meaning that they do not provide information about the chemistry that is occurring during reaction, yet they have been shown to accurately predict the results from thermal depolymerization for complex waste mixtures, such as food waste.[30−32] A handful of machine-learned models have been developed specifically for plastics, especially in the classification of polymer types.[33−35] The handful of studies that have used machine learning to predict depolymerization reactions have focused on single plastics and not mixtures[36] or understanding properties of single phases.[37] The few studies that have looked at mixed plastic streams have focused on specific and narrow ranges of operating conditions like temperature[38] or specific reactor types[39] and have utilized very little data to train and validate the models (≪100 data points).

The pertinent literature contains many hundreds of data points on plastic pyrolysis, and thus utilizing all of the published literature data can improve the accuracy of machine-learned models as well as expand the range of conditions and reactor types that can be studied. Naturally, utilizing the full range of published data requires an approach that can handle differences in feed, reaction conditions, reactor type, and the presence of catalyst. Fortunately, modern regression techniques have established ways to handle both numerical data as well as categorical data, such as reactor type or presence of catalyst.[40] Emerging methods for filling gaps in the values of dependent values—which are not uniformly reported—offer a way to maximize available data and harmonize between different reporting standards. The K-nearest neighbor (KNN) method has proven especially versatile for data harmonization in other fields but has yet to be applied in a plastic pyrolysis case.[41] Using models and methods such as these allows for efficient analysis of real waste streams and identification of promising experiments and conditions for future work.

The objective of this study is the development of machine-learned models for prediction of oil yields obtained by pyrolysis of common plastics and their mixtures over a wide range of conditions. Unlike previously reported studies, the current effort attempts to harmonize pyrolysis data arising from many different polymers—all of the "big six"—and for pyrolysis in different reactor types with and without the use of catalysts. The performance of seven different machine learning methods was comprehensively evaluated for accuracy. In particular, each of these models was trained, optimized, and

validated using a data set curated from the open literature encompassing a wide range of possible reaction conditions. The accuracy of model predictions was assessed with a test set that was set aside from the data set prior to training. The most accurate of these seven models was then used to predict the oil yields obtainable from pyrolysis of real plastic waste streams representative of U.S. Municipal Recycling Facilities (MRFs) and present in the Rhine River, a river with a high and well-characterized plastic load.[42,43] As a further example, the utility of the model was evaluated for making thermodynamic feasibility predictions of plastic pyrolysis. Finally, the machine-learned models and the proposed approach to develop them could become useful evaluative tools for advancing plastic pyrolysis as part of a comprehensive strategy to reduce plastic waste.

The paper is organized as follows: Section 2 encompasses the conceptual, methodological, and computational aspects of the proposed modeling and simulation framework. The study's main results are presented in Section 3 followed by a pertinent discussion. Finally, a few concluding remarks are provided in Section 4.

2. METHODS

2.1. Overview

The aim of this work is to develop a rigorous method for creating, validating, and applying machine-learned models to predict the yields that can be obtained from pyrolysis of waste plastics and their mixtures. Figure 1 shows the overall approach used for model
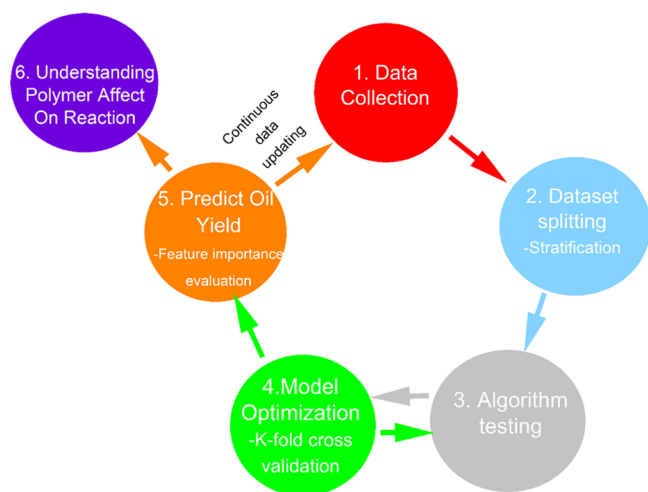

Figure 1. Outline of model development and deployment, including validation process for building and validating a robust model capable of predicting plastic pyrolysis oil yields.

development, refinement, selection, and application. The most important step is generation of the data set itself. Specifically, the data set was generated by retrieval of data published in the open literature with attempts to be inclusive up until the start of 2022. Once the data were collected and prepreparation steps were completed, seven different machine learning algorithms were tested and then validated using K-fold cross validation. These models were then used for predicting oil yields and understanding how polymer type in the feed affects the oil yields. Future sections will describe each of these steps in more detail.

2.2. Data Collection and Preparation

The data set was collected from the open literature and consists of 325 individual data points corresponding to 39 papers as seen in the

Supplementary Information Table SI.1. Studies describing pyrolysis of any of the "big six" plastics (PS, HDPE and LDPE, PP, PVC, and PET) were included in the retrieval effort with pyrolysis defined as experiments performed in different reactor types with and without a catalyst. The data were collected from reliable sources as described in pertinent peer-reviewed journal publications. In particular, the data set was derived from studies of both pure plastics and plastic mixtures and the distribution of data points corresponding to different types of pure plastics and mixed plastic streams is shown in Figure 2a. Of the plastics considered here, HDPE and LDPE along with PP are the most abundant in the current data set.

Studies on mixed plastic streams account for about one-third of the entire data set. Figure 2b shows the distribution of the mixed plastic data in terms of the number of plastics in the mixture. Binary and ternary mixtures are the most commonly studied mixtures, but several studies of mixtures, including all of the "big six" plastics, appear in the data set as well. For studies involving waste mixtures, the mass fraction in percent of each plastic type in the feed became its own independent variable input to the model. For pure streams, the mass fraction of the plastic being pyrolyzed was set to 100, while all other mass fractions were set to 0.0.

Along with the concentration of each of the plastics present in the feed, six other reaction parameters were also recorded in the data set and used as independent variables (called features): pyrolysis reaction temperature, heating rate, particle size, amount of feed utilized, catalyst, and reactor type. Pyrolysis reaction temperature, heating rate, particle size, and amount of feed utilized are all numerical and can be handled using typical regression methods. After careful deliberation, reaction time was not included as an independent variable as it is not well defined for different reactor types. More detail is provided in Section 3.

Unlike the quantitative variables, catalyst and reactor type are categorical variables, which require special treatment. In order for the model to be able to handle these categorical variables, they were one-hot encoded.[40] For example, catalyst was used as an independent variable with its value set to either 0 or 1 to represent the absence and presence of a catalyst, respectively. The model includes no information on the type of catalyst. Similarly, the reactor type was divided into five categories (batch, fixed bed, fluidized bed, horizontal tube, and semibatch) and each of these was described by its own independent one-hot encoded variable.

One potential drawback of using data from many different papers is inconsistencies in data reporting from lab to lab. The upshoot of data reporting inconsistencies is data gaps. To avoid this problem for dependent variables, pyrolysis oil yield was selected as the sole dependent variable as all studies of plastic pyrolysis report gravimetric oil yield. Oil yield was taken as the mass of liquid products recovered at room temperature and pressure. The phase behavior of the oil mixture depends on composition, and as a rough guide, it extends from products that are less volatile than pentane to compounds that melt at temperatures less than octadecane. Nonideal phase behavior and solvation effects mean that this range can only be used as a guideline. Since oil composition varies from study to study and is not uniformly reported in the literature and since oil recovery protocol also vary from study to study, reported oil yield data display natural variability that influences model predictability. Data gaps in the independent variables are unavoidable as not all studies report the six operating parameters selected for this study. Two common ways to handle data gaps are often used, omitting the data points that have missing data or using a data imputation algorithm such as KNN to fill in the data gaps.[44] Although omitting data with incomplete independent variable reporting is the more conservative of these approaches, preliminary tests found that eliminating even a handful of data points significantly and negatively impacted model performance as seen in the Supplementary Information. Therefore, this work used the KNN method to fill in the above data gaps. The KNN algorithm works by looking at K of the nearest data points to that which is unknown and making an educated prediction based on the votes of each of these points. K can be chosen in a multitude of ways, but the conventional approach is to use the square root of the number of data
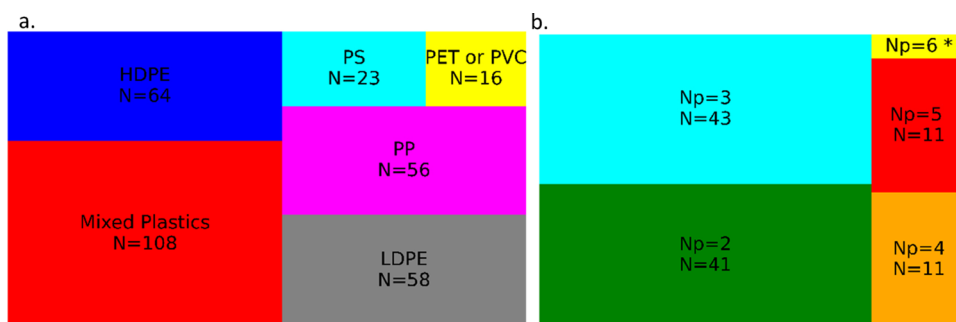
Figure 2. Visual representation of the 325 data point pyrolysis data set, showing (a) distribution between pure plastic studies and mixed plastic studies where *N* is the number of data points and (b) number of plastics in the mixed plastic studies where *Np* is the number of plastics present in the mixture. *Np* = 6 corresponds to two data points.

points:[45] for this data set, K was 17 for the training set size of 292 data points. KNN was needed to fill gaps for heating rate, particle size, and feed size. It should be pointed out that improved consistency of data reporting in future studies could potentially reduce the need for methods to handle data gaps.

2.3. Model Evaluation Metrics

Within the context of the present study, the mean absolute error (MAE) criterion was used to evaluate model accuracy. Eq 1 shows how MAE is calculated:

$$MAE = \frac{\sum_{i=1}^{N} |x_i - \hat{x}_i|}{N} \tag{1}$$

where N is the number of data points, x_i is the predicted value of x, and $\hat{x}_i$ is the experimental (true) value of x. Compared to other potentially viable alternatives, such as the root-mean square error and mean relative error criteria, MAE is the most widely used and easiest to interpret. The units of MAE in this work correspond to the units of oil yield (wt %).

2.4. Machine Learning Models

Seven different machine learning algorithms were evaluated in this study. Of the seven, three were linear methods and four were nonlinear methods. The linear methods are linear regression (LR), linear regression with lasso regression (LR-lasso), and linear regression with ridge regression (LR-ridge). Both lasso and ridge regressions are forms of LR that use penalty functions to shrink the model to fewer parameters.[29] The nonlinear models that are studied are decision tree (DT), eXtreme Gradient Boosting (XGBoost), random forest (RF), and artificial neural networks (ANN). DT, XGBoost, and RF are all tree methods[29] with DT using a single tree, while XGBoost and RF methods rely on a series of trees to improve accuracy.

Each of the models that were tested was run and optimized using Python 3.6[46] and scikit-learn 1.1.0 packages.[47] Model optimization involves tuning internal parameters, termed hyper parameters. The optimal hyper parameters for each model can be found in the Supporting Information Table SI.2. For all methods other than the tree models, the data had to be normalized due to the large difference in the values for the different reaction parameters (i.e. particle size ranges from microns to mm and feed sizes from mg to thousands of g). When normalization was required, a *Z* transform was performed on the data, including all of the model features. Eq 2 shows how the *Z* transform of a variable is calculated:

$$Z_i = \frac{x_{i,j} - \bar{x}_i}{\sigma_i} \tag{2}$$

where Z_i is the new value, $\bar{x}_i$ is the mean of feature i, and σ_i is the standard deviation of feature i. DT methods by nature are not sensitive to these differences, and unscaled data were used for these model types.[29]

2.5. Data Splitting

Development of rigorous, predictive machine-learned models requires training the model on a data set followed by testing its accuracy for prediction of a separate data set. These two data sets are called the training and testing sets. For the current study, 10% of the data were removed prior to model optimization, placed into a "vault," and not touched again until all of the models were optimized to be able to test the models on "new" data that were not "seen" during model development. The remaining 90% of the data were used for KNN and model training. Dividing data into training and test sets is common practice to prevent over fitting; inadvertent use of test data for model optimization is termed as "data snooping" and can reasonably be expected to result in over fitting.[29] The splitting of the data set into training and test sets is done through random selection.

Model optimization on the training data utilized a method of validation, which further separates the training data into separate subsets. The model validation method selected is termed as K-fold cross validation, which separates the training data into K subsets. Here, K was set equal to 10, meaning that the training data were split into 10 equal groups. Nine of these groups were used to train the model (training data set) and one was used to validate (validation data set). This was repeated sequentially until all 10 groups had acted as the validation set. Cheng et al. provided a visual representation of how the K-fold method works.[30] K-fold cross validation helps to prevent overfitting or fortuitus fitting as the model must be trained on all of the data, and the final result represents an average of all 10 models.

One of the potential pitfalls of working with small data is ensuring representative splitting into test and training data sets. For example, relatively few studies are published that include low-yield data, which means that random splitting into test and training sets may result in a training set that completely lacks low-yield data. The result is a model that cannot accurately predict data with low yields. To prevent the problem of nonrepresentative data splitting, the data set was stratified based on yield prior to splitting. Stratification grouped data by reported oil yield to create a set of bins each with an equal number of data points (10 here). During splitting, at least one data point from each bin was included in the test set to ensure that it was representative of the full range of data present in the training set. The data were then rerandomized prior to training and validation.

3. RESULTS/DISCUSSION

The goal of this study was to use published data to develop a data-driven, machine-learned model to predict oil yields obtained from plastic pyrolysis. A data-driven model has the advantage that its predictions do not require detailed understanding of chemical mechanisms. Instead, reaction conditions can be used as the independent variables or features and inputs to a regression-type model that does not require knowledge about the underlying chemistry.

In addition to pyrolysis, the kindred method of HTL has been reported for conversion of waste plastics to useful oils.[48] A universal tool that can predict HTL or pyrolysis yields obtainable from waste plastics could be very useful. Unfortunately, most studies on HTL have focused on plastics other than those in common commercial use, meaning that the HTL and pyrolysis data sets have minimal overlap and rendering simultaneous modeling of HTL and pyrolysis impractical at present.

After initial tests, reaction time was not included as an independent variable due to the fact that reaction time is a fundamentally different concept depending on the reactor type. For example, reaction time is well defined as the residence time in a closed, batch reactor on the one extreme and in a plug-flow continuous reactor on the other. For the most common pyrolysis reactors, including spouted bed reactors, swept-batch or semicontinuous reactors, and fluidized bed reactors, residence time of the plastic within the reactor is either poorly defined or impossible to compare with other types of reactors. Excluding certain reactor types limits the amount of data available for the model. Since the model is already pushing the lower boundary of how much data is required for robust predictions, all reactor types were included as independent variables and reaction time was excluded. In fact, preliminary tests (seen in the Supplementary Information) that conflated the different types or reaction times did not provide accurate predictions of pyrolysis oil yield.

Many different types of machine-learned models have been published in the literature for a range of different engineering applications.[30−39] Accordingly, a wide variety of model types were evaluated as the ideal model for a given data set cannot be predicted a priori. The seven models studied included three linear models (LR, ridge regression, and lasso regression), three nonlinear tree models (DT, RF, and Xtreme Gradient Boosting), and one nonlinear model (ANN). These seven models were chosen as they are representative of successful model types previously studied in the literature for similar problems.[30−32] LR represents the simplest starting point and is a good basis of comparison. Lasso and ridge regressions are well-known modifications of the standard LR model that add an error function intended to improve accuracy.[29] DT is the simplest Boolean model, and RF and XGBoost in particular were chosen because their use of multiple trees as well as subsets of the data improves their accuracy, especially when working with small data sets.[49]

Each of the seven models consists of multiple parameters (termed as "hyper parameters" in the machine learning literature[50]), which the user must select based on the application at hand. Hyper parameter values are optimized during cross validation to minimize model error. The optimal model hyper parameters for each individual model can be found in the Supplementary Information Table SI.2.

After model optimization, prediction accuracy was assessed for both the validation and test sets. MAE was used as the error metric in this work. The most accurate model should minimize MAE, exhibit minimal change between the validation and test set MAEs, and have low standard deviation arising from validation. Each of these plays an important role. MAE of the test set is the truest measure of the predictive accuracy of the model. Agreement between the test set and validation MAEs, as well as minimal validation standard deviation, guards against over fitting and fortuitous division between validation and test sets. Using validation set MAE as the sole basis for model

selection therefore risks model overfitting, which detracts from accuracy of predictions for new data not included in the original data set.

Figure 3 shows both the validation set and test set MAE for all seven models, as well as the standard deviation of the
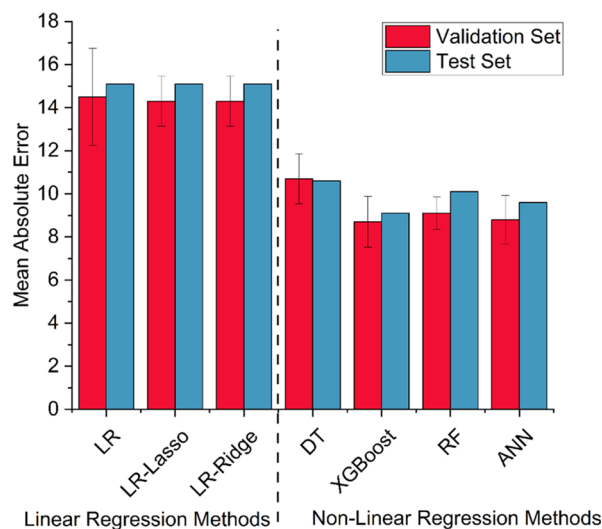

Figure 3. Mean absolute error (MAE) of the validation and test sets for linear regression (LR), lasso regression (LR-lasso), ridge regression (LR-ridge), decision tree, eXtreme Gradient Boosting (XGBoost), random forest (RF), and artificial neural network (ANN).

validation set error. The MAE of the benchmark LR model is approximately 14 for both validation and test set analysis. The corresponding standard deviation of the validation set MAE is nearly 3. While the agreement between validation set and test set MAE is reassuring, the absolute value of these MAEs is not sufficient for most applications and the large standard deviation of the validation set MAE suggests that the regression is prone to errors due to fortuitous data selection. The lasso and ridge modifications of LR offer no benefit, indicating that the addition of an error function is insufficient to capture what is inherently a highly nonlinear data set. Accordingly, linear models should not be expected to provide accurate pyrolysis yield predictions, except possibly over very narrow ranges of conditions where the problem might be nearly linear.

All of the nonlinear models tested here provided superior performance compared to the linear models. Of the four nonlinear methods, the MAE provided by the RF, XGBoost, and ANN models is less than that provided by the DT model. While the DT model fits the validation data better than the linear methods, the MAE of the validation and test sets along with the validation set standard deviation of DT is greater than that of the other three nonlinear models. In comparison, the modest value of validation set standard deviation observed for RF indicates that it is much less sensitive to the subset of data it is trained on than any of the other models tested here. The comparison of validation set standard deviations recommends selection of RF over ANN or XGBoost, especially since all three methods yield similar values of MAE. The performance of RF, ANN, and XGBoost is nearly identical, and therefore, any method could justifiably be used for further analysis. In this work, XGBoost was chosen for future model applications due to its relative simplicity, low MAE for validation and test sets (8.7 and 9.1%, respectively), and low standard deviation in

validation (±1.2%). A parity plot of the experimental vs predicted oil yields of the test set for the optimal XGBoost model can be found in the Supplementary Information Figure SI.1. The majority (67%) of predictions fall within ±10% of the experimentally reported values, with the remainder contributing to the observed value of MAE (9.1%) as seen in Table SI.3.

The MAE of the XGBoost model captures the residual error of the current data set as it was extracted from the literature. Residual error likely arises from several factors, and reducing the error can be achieved in several ways. First, the available data set (<500 data points) is at the lower limit of what can be studied using machine learning. Increasing the amount of data—for example, by updating the model periodically as new data appear in the literature—should improve accuracy over time.[51] Second, the experimental data themselves are subjected to experimental uncertainty, which is on the order of several percent. Experimental uncertainty detracts from model accuracy, and reducing the inherent experimental error should reduce the MAE that the model can achieve. Experimentalists should adopt methods such as mass balance and carbon balance closure to ensure data quality and reduce uncertainty. Third, different studies use slightly different methods—including reactor types, as previously discussed, and also analytical methods for recovering and quantifying oil yields. Differences in methodology give rise to systematic differences between oil yields reported in different studies. Greater consistency in analytical methods, especially in recovery conditions and oil yield definitions (e.g., instead of reporting total gravimetric yield and reporting yield obtained over a certain distillation range), should improve consistency in the data set and help to reduce uncertainty. Application of XGBoost-based regression on a single, uniform data set—of sufficient size (>500 data points)—can realistically be expected to achieve MAE values less than seen here. That data set does not exist in the public domain at this time.

A common analysis of regression error is to determine the statistical significance of the various features. For linear models, a common method is the F test.[52] For DT models, the corresponding metric is feature importance. Figure SI.2 provides the feature importance extracted from the XGBoost model. Plastic type and particle size have similar levels of importance (~10%), indicating that they have statistically significant effects on model predictions. Interestingly, the feature importance of catalyst was negligible compared with the other features. Effectively, the pyrolysis oil yields obtained using a catalyst may often be reproduced by increasing the pyrolysis temperature. However, other studies found that catalyst use has minimal or even negative effects on yields and instead mainly effects product distribution[53]—a dependent variable not considered here, outside of how product distribution affects oil yield. Data on the use of catalysts may be biased to lower temperatures that accentuate the perceived benefit on yield. In these cases, the benefit of using a catalyst is not so much as to increase oil yields but to lower the temperature at which an acceptable yield can be obtained. The net effect of catalyst use in these cases is to improve energy yield, an outcome that is not explicitly predicted by the models presented here.

The feature importance also highlights that the influence on oil yield of pyrolysis of plastics falls into two categories: factors that capture chemical reactivity and factors that are related to heat transfer. The chemical reactivity factors include plastic type, temperature, and presence of catalyst, and together they account for 57.2% of the observed correlation. Heat transfer factors include particle size, plastics loading, and heating rate, which combine to account for 40% of the correlation. The importance of chemical reactivity and heat transfer factors is consistent with intuition,[54] which is a comforting result for a data-driven model that is ignorant of physics.

Of all of the features considered in the present study, the value of the feature importance assigned to temperature was the greatest (43.6%), indicating that temperature had the most significant effect on pyrolysis yields. Due to the high feature importance of temperatures, understanding how temperature appears in the current data set is critical to understand before the corresponding model can be used to make predictions of new systems. Since model predictions can only be trusted for conditions that fall within the limits established by the data set, the temperature limits where model predictions can be trusted are based on the appearance of temperature in the data set. As a temperature-activated process, pyrolysis temperatures will naturally be biased to values that result in appreciable oil yields. Researchers use their knowledge of the system to determine the temperatures at which they run their experiments, meaning that temperatures much less and much greater than optimal do not appear in the literature. The result is that predictions for conditions that fall outside of the limits appearing in the data set cannot be trusted. Because each plastic has its own temperature, which optimizes pyrolysis oil yields, the available data set must be evaluated to understand the temperature range over which each plastic has been studied.

Figure 4 plots the temperature range covered by each of the plastics included in this study, as well as the temperature range covered by pyrolysis of mixtures, as a box−line plot where the box represents the 25 and 75% limits of the available data and the line represents the absolute limits. As expected, the temperature range studied for each plastic varies with the plastic, with PP, PS, and PET tending to be studied at low
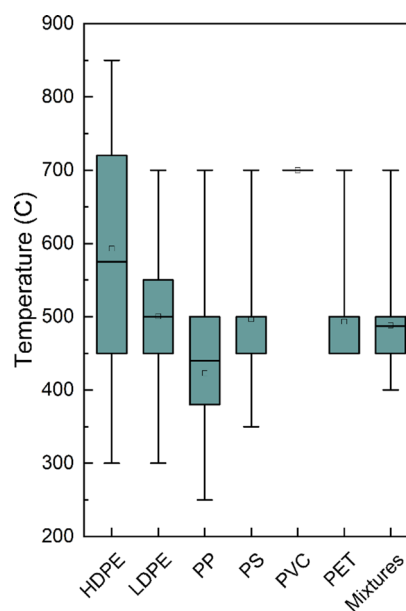

Figure 4. Temperature distribution for each polymer and mixtures in the data set showing the 25−75% range (box), the min−max range (line), the median line, and the mean (white small square).

Table 1. Comparison of Plastics in European and American Municipal Solid Waste and Their Predicted Oil Yields

location	HDPE (%)	LDPE (%)	PP (%)	PS (%)	PVC (%)	PET (%)	predicted oil yield (%)
USA[15]	17.86	24.08	22.78	6.29	2.4	14.79	35.3
Europe[20]	44.4	0	21.2	13.3	12.2	8.9	48.9
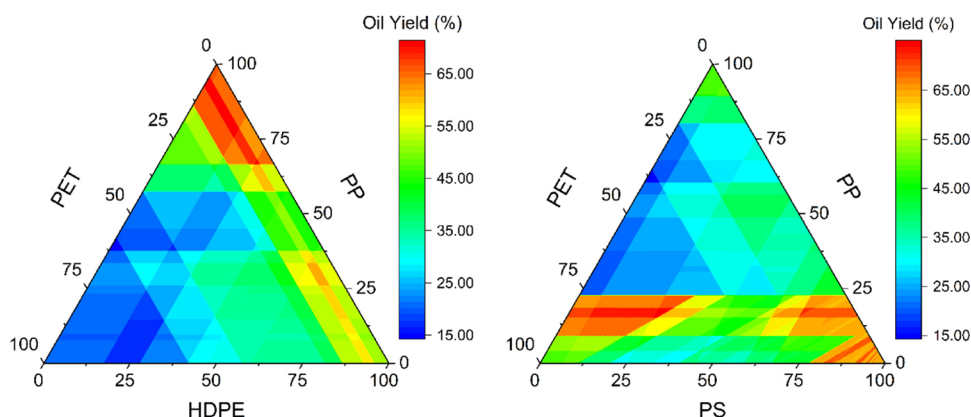
Figure 5. Predicted oil yields for mixtures of HDPE, PP, and PET and PS, PP, and PET pyrolyzed in a horizontal tube reactor at 500 °C with a heating rate of 10 °C/min, a particle size of 13 mm, and plastics loading of 200 g without catalyst.

temperatures and PE at higher ones. PVC pyrolysis has been studied at a single temperature (700 °C). Fortunately, in aggregate, the temperature range over which pyrolysis yields are optimized from 400 to 700 °C is well covered in the data set—provided that pure PVC is excluded from the analysis. PVC is generally regarded as an unsuitable feed for plastic pyrolysis due to the formation of hydrochloric acid during pyrolysis. Hydrochloric acid is corrosive, and chlorine contamination detracts from fuel properties.[55] Hence, a predictive model that is appropriate for pure or mixed plastics with the exception of PVC in the range from 400 to 700 °C is valuable for many applications. Predictions made outside this range are not physically realistic with the consistent over prediction of pyrolysis oil yields at temperatures less than 400 °C.

Figures 3 and 4 indicate that an XGBoost model can make accurate pyrolysis oil yield predictions in the temperature range from 400 to 700 °C. The next step was to demonstrate the method for two realistic applications: pyrolysis of plastics collected at MRFs and pyrolysis of plastics present in river outflows to the ocean.

A major benefit of pyrolysis colocated with MRFs is that they exist in centralized locations where plastic is collected. Although the exact composition of plastics at these facilities can change from location-to-location, representative averages exist for different countries/areas. These averages can vary significantly from country to country, and therefore, the expected oil yields from converting the plastic to fuels are also expected to vary. The machine-learned model can predict oil yields for these different streams. To ensure reliable model predictions, the model was run in a batch reactor and at a constant temperature of 500 °C well within the limits suggested by Figure 4.

Table 1 provides average compositions and corresponding predictions of oil yields for municipal solid waste (MSW) collected in both the USA and EU. The two plastic streams vary considerably, especially in their relative ratios of HDPE and LDPE and their amounts of PS and PVC. These composition differences translate into predicted oil yields that differ by almost 15% from one another, with the oil yield

predicted for pyrolysis of EU plastic waste being greater than that predicted for U.S. waste. That difference is likely related to the greater PS content of EU waste than U.S. waste since PS is more easily pyrolyzed to obtain high oil yields than PP or PE.[56]

The XGBoost model can also be used to explore the benefits of partial separation for optimizing oil yields. In addition to PVC, oil yields obtained from pyrolysis of PET are generally less than those obtained from HDPE, LDPE, PS, and PP.[56,57] Consistent with this expectation, the XGBoost model predicts that pyrolysis of pure PET at 500 °C results in 28% oil yield much less than that predicted for PP, PS, or HDPE at the same conditions (74, 70, and 63%, respectively, as seen in Supplementary Information Figure SI.3). Similarly, whereas HDPE, LDPE, PS, and PP are all hydrocarbon plastics that yield oils with properties similar to hydrocarbon fuels, PET is an oxygenate that yields small oxygenated compounds, especially ethylene groups and benzoic acid,[58] that are more suitable for use as monomers than as fuels. Finally, of the six commonly used plastics, PET has the most robust recycling market, meaning that recycled PET has an existing valorization channel that the other plastics lack.[59] For these reasons, separation of PET prior to pyrolysis is worth evaluating. Both new and existing[60] technologies can be used for PET separation.

To evaluate the effect of PET on pyrolysis oil yields, the XGBoost model was run many times for different ratios of PET in a PP and HDPE mixture and separately for PET in a PP and PS mixture. The combinations of plastics shown in Figure 5 were chosen due to the prevalence of HDPE, PP, and PS in both the USA and European MSW.[15,20] The result of these simulations is ternary diagrams relating composition to predicted oil yields as shown in Figure 5. The oil yield analysis was performed for a mixture of LDPE, PP, and PET and results can be found in the Supplementary Information as Figure SI.3. The general trend observed in Figure 5 is that oil yields decrease with increasing PET content, strongly recommending PET removal prior to pyrolysis. Some interesting areas of potential synergy exist, for example mixtures of approximately 80% PET and 20% PP.

In addition to waste already collected at MRFs, significant waste ends up in the environment. Waste plastic in rivers throughout the world is especially damaging as it not only pollutes the river ecosystems but also eventually pollutes the ocean as rivers act like highways to transport waste plastic to the ocean.[9] Rivers polluted with waste plastic exist all over the world,[11] and targeting these rivers as sources of plastic for conversion to oil could cut off these plastic highways into the world's oceans. The complete system would include a mechanism to collect the plastic, remove it from the river, feed it to the pyrolysis process, pyrolyze the plastic, and recover the oil similar to a system designed for ocean plastics.[61]

The composition of plastics varies significantly from river to river[62] due to differences in regional industry and community usage. The natural consequence of these differences in composition is differences in oil yields obtainable by pyrolysis. Prioritizing specific rivers to implement plastic capture and conversion systems allows for optimal use of finite resources. One way to prioritize rivers is to predict the oil yields that can be obtained for pyrolyzing the plastic present in them, potentially feeding those predictions into economic models to optimize return on investment. Alternatively, prioritization could be based on relative ecological importance (i.e., proximity to fragile ecosystems), in which case predictions of oil yields are necessary to build a business model.

As a concrete example, the XGBoost model was used to predict the oil yields obtainable from pyrolysis of plastic found in the Rhine River in Europe. This river was chosen for the abundantly available data about plastic types and concentrations found along the length of the river.[42,43] The Rhine River waste plastic consists of mainly HDPE, PP, and PS (27, 37, and 26% by mass respectively). This average composition was then used along with a common set of reaction parameters (found in the Supplementary Information Tables SI.4 and SI.5) to predict an oil yield range of 44.1 to 56.7% for pyrolysis temperatures from 400 to 650 °C comparable to the value observed for the MRF present in the EU. The predicted oil yields for each temperature can be found in the Supplementary Information Table SI.6.

The predicted oil yields were then inputted to a stochastic process model that utilizes the Monte Carlo (MC) technique for evaluating thermodynamic outcomes[61] of a plastic recovery and conversion process as a function of pyrolysis temperature and feed rate. In addition to the pyrolizer itself, the process includes pumps, shredders, blowers, filters, and other peripheral equipment. Modifications to the model previously published by Belden et al. can be found in the SI. In brief, the system consists of collection, shredding, pyrolysis, and oil recovery steps. The energy requirements for all auxiliary steps were taken from manufacturer specifications and assumed not to allow turn down to less than full power consumption. The energy requirement of the pyrolysis step was taken from thermochemical analysis of the enthalpy of reaction.

A major unknown in the analysis is the amount of plastic exiting the river that can be harvested. Accordingly, simulations were performed over a range of plastic feed rates with the goal of identifying under what conditions the process can be expected to be self-sufficient. All equipment in the process were off-the-shelf parts, and their energy consumption was based on vendor specifications and did not scale with plastic feed rate. The exergy required for the pyrolysis reactor was based on the heat of reaction, and the exergy of the oil product was the process output; both of these values scaled linearly

with the mass flow rate of plastic entering the process. To handle parameter uncertainty, the pyrolysis oil yield was handled as a stochastic variable in the MC simulation; its value was varied to reflect the observed model MAE (±9.1%). The remaining stochastic variables can be found in the Supplementary Information Table SI.6.

Figure 6 shows the results of this analysis as a function of flow rate over a range of realistic temperatures as the
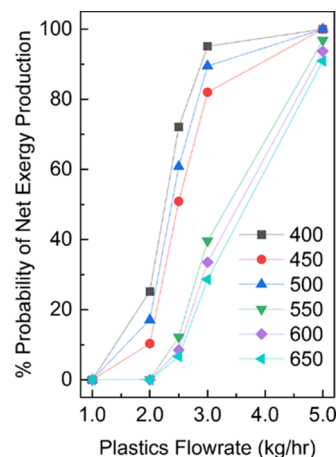

Figure 6. Percent probability of net exergy production for the Rhine River from 1 to 5 kg/h flowrates and over a pyrolysis temperature range of 400 to 650 °C.

probability of producing more exergy (based on fuel heating value) than it consumed ("net exergy production"). For feed rates of 1.0 and 2.0 kg/h, the process is unlikely to produce more exergy than it consumes (<20%) since the energy required by the peripheral equipment is predicted to be greater than that in the fuel product. Between a feed rate of 2.0 and 2.5 kg/h and for temperatures less than 550 °C, the probability of net exergy production increases by a factor of 4; this sharp increase corresponds to the point at which the exergy embodied in the pyrolysis oil exceeds that required by the auxiliary equipment (~5 kW of fixed consumption). Increasing from 2.5 to 3.0 kg/h results in >80% probability of net exergy production for pyrolysis temperatures less than 550 °C.

In general, the model predicts that thermodynamic efficiency decreases with increasing pyrolysis temperature, a consequence of the high reactivity of the Rhine River plastic mixture. The exception to this trend is that the probability of net exergy production predicted at 500 °C is greater than that predicted at 450 °C. Here, the model predicts that the oil yield at 450 °C is slightly less than at 500 °C (53% compared to 55.2%), accounting for the counter-intuitive finding. In fact, this difference in oil yield is within model uncertainty, meaning that the model indicates operating in the range between 400 and 500 °C for optimal thermodynamic efficiency. This range can be then used to minimize time spent in the evaluation of process performance with actual plastic mixtures.

Figure 6 can be interpreted to predict that pyrolytic conversion of Rhine River plastic is very likely to be energy self-sufficient, assuming that 100% of the plastic carried by the river can be captured. An estimated 20 to 30 tons of plastic is discharged from the Rhine River annually,[63] which corresponds to plastic flowrates of ~2 to 3 kg/h, right in line with the predictions seen here for thermodynamically self-contained capture and pyrolytic conversion. Lower ends of this plastic

exit rate may argue for the accumulation of plastic and the semicontinuous operation of the conversion system to permit modest process scale-up to achieve energy self-sufficiency—or for redesign to improve efficiency. The analysis shown in Figure 6 assumes fixed power requirements for peripheral equipment, whereas the energy required for pyrolysis is embodied in the fuel scale with plastic feed rate. If the energy consumed by peripheral equipment can somehow be scaled to the feed rate—i.e., less energy might be consumed by the shredder for a dilute mixture than a concentrated one, then more favorable outcomes can be achieved for lower flow rates than that are shown here. The analysis presented here for the Rhine River also suggests that other rivers with higher plastic discharge rates, including the Yangtze, Ganges, and Xi,[11] may yield self-sustaining plastic removal and pyrolysis systems, especially when complete capture of plastics is not possible, even accounting for the possibility that the oil yields for those rivers may be less favorable than predicted for the Rhine River.

The examples of MRF and Rhine River plastic pyrolysis demonstrate how the current model can be used. The emphasis has been placed on predicting pyrolysis oil yields, with the end application of replacing petroleum-derived fuels. Because the origin of synthetic plastics is petroleum, combustion of plastic-derived fuels releases the same amount of CO_2 into the atmosphere as does combustion of petroleum-derived fuels. Using estimates of the global use of plastics[64] minus PVC and PET and the current model, approximately 130 million metric tons of plastic-derived fuel can be produced annually. Using typical emissions factors,[65] combustion of this plastic-derived fuel would be equivalent to 230 million tons/year of CO_2, approximately 1% of annual emissions.[66] That stated, on a per mass basis, the environmental impact of plastic entering the environment is almost certainly greater than that of CO_2 despite the threat of climate change.

Nonetheless, the CO_2 emissions analysis indicates that using pyrolysis to produce fuels incurs unacceptable climate penalties. Therefore, pyrolysis to fuel should be used as part of a comprehensive strategy, which includes pyrolysis for chemical production, increased recycling, and increased use of biodegradable plastics, to prevent plastic entering landfills or the environment. Similarly, the climate impact of using plastic-derived fuels can be minimized if they are reserved for use in applications that are difficult to decarbonize using available technologies, such as shipping or air transport.[67] Alternatively, to avoid new CO_2 emissions entirely, pyrolysis can be used to produce products other than fuels, such as lubricants, monomers, or other chemicals. The current model method can easily be adapted for these alternative products provided that data are available.

The results presented here indicate that the accuracy of the current model is sufficient for many practical applications. More generally, the accuracy of the model with the available data set indicates substantial scope for improvement as more data become available. Some of the uncertainty of the current model likely arises from incomplete reporting of independent variables—necessitating the use of KNN to fill data gaps—and a lack of a uniform way to report oil yields. A consistently used volatility-defined measure of oil yield based on simulated distillation or thermogravimetric analysis of the product would improve data quality and model predictability. More uniform and quantitative reporting of individual product yields would allow the model to be adapted for predictions of chemicals production.

4. CONCLUSIONS

Seven machine-learned models were optimized and evaluated for predicting oil yields obtained from pyrolysis of the big six plastics using feed composition, reaction conditions, reactor type, and the presence or absence of catalyst as independent variables. Of these seven models, the accuracy profile of nonlinear models was found to be superior to the one associated with conventional linear models included in the study, with XGBoost providing the most accurate predictions based on the MAE criterion. Because of the content of the specific data set used to train the proposed machine-learned models, they are expected to be accurate in the temperature range from 400 to 700 °C, inclusive.

As a demonstration, the XGBoost model was used to predict the oil yields expected from pyrolysis of the U.S. and EU plastic recycling waste, with the finding that EU waste appears to be a superior feed candidate for pyrolysis. The same regression model was used as part of a probabilistic Monte Carlo-based thermodynamic analysis that was performed on a process to remove plastic from the Rhine River, shred it, feed it to a pyrolyzer, and convert it into fuel. Thermodynamic analysis indicated that the process could be a net exergy producer under realistic conditions, provided that the scale of the pyrolyzer was sufficient to produce enough oil to offset fixed energy requirements associated with the peripheral equipment.

The model used here is therefore deemed sufficient for many practical applications involving waste plastic pyrolysis and in particular can be used in conjunction with thermodynamic analysis to evaluate feasibility. As more data are reported, the model can gradually be refined and its predictive and performance evaluative capacity considerably enhanced. In addition to pyrolysis, the general methodological, modeling, and simulation framework proposed in the present study could be readily applied to different technology options as well as inform strategies and policy responses for plastic waste reduction and management.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsengineeringau.2c00038.

> Supporting Information contains: data set data table; initial testing results; model parameters; supporting model results and figures; and thermodynamic model modifications (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Michael T. Timko — *Department of Chemical Engineering, Worcester Polytechnic Institute, Worcester, Massachusetts 01609, United States;* ⓘ orcid.org/0000-0001-8767-1613; Email: mttimko@wpi.edu

Authors

Elizabeth R. Belden — *Department of Chemical Engineering, Worcester Polytechnic Institute, Worcester, Massachusetts 01609, United States*

Matthew Rando — *Department of Chemical Engineering, Worcester Polytechnic Institute, Worcester, Massachusetts 01609, United States*

Owen G. Ferrara − *Department of Chemical Engineering, Worcester Polytechnic Institute, Worcester, Massachusetts 01609, United States*

Eric T. Himebaugh − *Department of Chemical Engineering, Worcester Polytechnic Institute, Worcester, Massachusetts 01609, United States*

Christopher A. Skangos − *Department of Chemical Engineering, Worcester Polytechnic Institute, Worcester, Massachusetts 01609, United States*

Nikolaos K. Kazantzis − *Department of Chemical Engineering, Worcester Polytechnic Institute, Worcester, Massachusetts 01609, United States*

Randy C. Paffenroth − *Department of Mathematical Sciences and Department of Computer Science, and Data Science Program, Worcester Polytechnic Institute, Worcester, Massachusetts 01609, United States*

Complete contact information is available at:
https://pubs.acs.org/10.1021/acsengineeringau.2c00038

Author Contributions

CRediT: **Elizabeth R Belden** conceptualization (equal), data curation (equal), formal analysis (equal), investigation (equal), methodology (equal); **Matthew Rando** conceptualization (equal), data curation (equal); **Owen G Ferrara** conceptualization (equal), data curation (equal); **Eric T Himebaugh** conceptualization (equal), data curation (equal); **Christopher A Skangos** conceptualization (equal), data curation (equal); **Nikolaos K. Kazantzis** conceptualization (equal), methodology (equal); **Randy C Paffenroth** conceptualization (equal), methodology (equal).

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

■ REFERENCES

(1) *Annual production of plastics worldwide from 1950 to 2020 (on millions metric tons)*. https://www.statista.com/statistics/282732/global-production-of-plastics-since-1950/ (accessed July 26, 2022).

(2) *Plastics: Material- Specific Data.* https://www.epa.gov/facts-and-figures-about-materials-waste-and-recycling/plastics-material-specific-data (accessed Jul 26, 2022).

(3) Chen, H. L.; Nath, T. K.; Chong, S.; Foo, V.; Gibbins, C.; Lechner, A. M. The plastic waste problem in Malaysia: management, recycling and disposal of local and global plastic waste. *SN Appl. Sci.* **2021**, *3*, 437.

(4) Cole, M.; Lindeque, P.; Fileman, E.; Halsband, C.; Goodhead, R.; Moger, J.; Galloway, T. S. Microplastic Ingestion by Zooplankton. *Environ. Sci. Technol.* **2013**, *47*, 6646−6655.

(5) Barboza, L. G. A.; Dick Vethaak, A.; Lavorante, B. R. B. O.; Lundebye, A.-K.; Guilhermino, L. Marine microplastic debris: An emerging issue for food security, food safety and human health. *Mar. Pollut. Bull.* **2018**, *133*, 336−348.

(6) Hopewell, J.; Dvorak, R.; Kosior, E. Plastics recycling: challenges and opportunities. *Philos. Trans. R. Soc., B* **2009**, *364*, 2115−2126.

(7) Filiciotto, L.; Rothenberg, G. Biodegradable Plastics: Standards, Policies, and Impacts. *ChemSusChem* **2021**, *14*, 56−72.

(8) Lebreton, L.; Slat, B.; Ferrari, F.; Sainte-Rose, B.; Aitken, J.; Marthouse, R.; Hajbane, S.; Cunsolo, S.; Schwarz, A.; Levivier, A.; Noble, K.; Debeljak, P.; Maral, H.; Schoeneich-Argent, R.; Brambini, R.; Reisser, J. Evidence That the Great Pacific Garbage Patch Is Rapidly Accumulating Plastic. *Sci. Rep.* **2018**, *8*, 4666.

(9) Schmidt, C.; Krauth, T.; Wagner, S. Export of Plastic Decris by Rivers into the Sea. *Environ. Sci. Technol.* **2017**, *51*, 12246−12253.

(10) Thompson, R. C.; Moore, C. J.; Saal, F. S. V.; Swan, S. H. Plastics, the Environment and Human Health: Current Consensus and Future Trends. *Philos. Trans. R. Soc., B* **2009**, *364*, 2153−2166.

(11) Lebreton, L.; Zwet, J. V. D.; Damsteeg, J.-W.; Slat, B.; Andrady, A.; Reisser, J. River plastic emissions to the world's oceans. *Nat. Commun.* **2017**, *8*, 16511.

(12) Al-Salem, S. M.; Lettieri, P.; Baeyens, J. Recycling and recovery routes of plastic solid waste (PSW): A review. *Waste Manage.* **2009**, *29*, 2625−2643.

(13) Joseph, B.; James, J.; Kalarikkal, N.; Thomas, S. Recycling of medical plastics. *Adv. Ind. Eng. Polym. Res.* **2021**, *4*, 199−208.

(14) Thiounn, T.; Smith, R. C. Advances and approaches for chemical recycling of plastic waste. *J. Polym. Sci.* **2020**, *58*, 1347−1364.

(15) *Advancing Sustainable Materials Management: Facts and Figures*, 2018. https://www.epa.gov/sites/default/files/2021-01/documents/2018_ff_fact_sheet_dec_2020_fnl_508.pdf (accessed December 2021).

(16) Geyer, R.; Jambeck, J. R.; Law, K. L. Production, use, and fate of all plastics ever made. *Sci. Adv.* **2017**, *3*, No. 1700782.

(17) Chen, H.; Wan, K.; Zhang, Y.; Wang, Y. Waste to Wealth: Chemical Recycling and Chemical Upcycling of Waste Plastics for a Great Future. *ChemSusChem* **2021**, *14*, 4123−4136.

(18) Ellis, L. D.; Rorrer, N. A.; Sullivan, K. P.; Otto, M.; McGeehan, J. E.; Román-Leshkov, Y.; Wierckx, N.; Beckham, G. T. Chemical and biological catalysis for plastics recycling and upcycling. *Nat. Catal.* **2021**, *4*, 539−556.

(19) Valero, A.; Valero, A. Thermodynamic Rarity and Recyclability of Raw Materials in the Energy Transition: The Need for an In-Spiral Economy. *Entropy* **2019**, *21*, 873.

(20) Williams, P. T.; Slaney, E. Analysis of products from the pyrolysis and liquefaction of single plastics and waste plastic mixtures. *Resour., Conserv. Recycl.* **2007**, *51*, 754−769.

(21) Qureshi, M. S.; Oasmaa, A.; Pihkola, H.; Deviatkin, I.; Tenhunen, A.; Mannila, J.; Minkkinen, H.; Pohjakallio, M.; Laine-Ylijoki, J. Pyrolysis of plastic waste: Opportunities and challenges. *J. Anal. Appl. Pyrolysis* **2020**, *152*, No. 104804.

(22) Elliott, D. C.; Biller, P.; Ross, A. B.; Schmidt, A. J.; Jones, S. B. Hydrothermal liquefaction of biomass: Developments from batch to continuous process. *Bioresour. Technol.* **2015**, *178*, 147−156.

(23) Anene, A. F.; Fredrksen, S. B.; Saetre, K. A.; Tokheim, L.-A. Experimental study of thermal and catalytic pyrolysis of plastic waste components. *Sustainability* **2018**, *10*, 3979.

(24) Zhao, P.; Yuan, Z.; Zhang, J.; Song, X.; Wang, C.; Guo, Q.; Ragauskas, A. J. Supercritcal water co-liquefaction of LLDPE and PP into oil: properties and synergy. *Sustainable Energy Fuels* **2021**, *5*, 575−583.

(25) Quesada, L.; Calero, M.; Martín-Lara, M. Á.; Pérez, A.; Blázquez, G. Production of an Alternative Fuel by Pyrolysis of Plastic Wastes Mixtures. *Energy Fuels* **2020**, *34*, 1781−1790.

(26) Harmon, R. E.; SriBala, G.; Broadbelt, L. J.; Burnham, A. K. Insight into Polyethylene and Polypropylene Pyrolysis: Global and Mechanistic Models. *Energy Fuels* **2021**, *35*, 6765−6775.

(27) Kruse, T. M.; Levine, S. E.; Wong, H.-W.; Duoss, E.; Lebovitz, A. H.; Torkelson, J. M.; Broadbelt, L. J. Binary mixture pyrolysis of polypropylene and polystyrene: A modeling and experimental study. *J. Anal. Appl. Pyrolysis* **2005**, *73*, 342−354.

(28) Kruse, M.; Sang Woo, O.; Broadbelt, J. Detailed mechanistic modeling of polymer degradation: application to polystyrene. *Chem. Eng. Sci.* **2001**, *56*, 971−979.

(29) James, G.; Witten, D.; Hastie, T.; Tibshirani, R., *An Introduction to Statistical Learning with Applications in R*; Springer: New York, 2013.

(30) Cheng, F.; Belden, E. R.; Li, W.; Shahabuddin, M.; Paffenroth, R. C.; Timko, M. T. Accuracy of predictions made by machine learned models for biocrude yields obtained from hydrothermal liquefaction of organic wastes. *Chem. Eng. J.* **2022**, *442*, No. 136013.

(31) Zhu, X.; Li, Y.; Wang, X. Machine learning prediction of biochar yield and carbon contents in biochar based on biomass characteristics and pyrolysis conditions. *Bioresour. Technol.* **2019**, *288*, No. 121527.

(32) Ascher, S.; Watson, I.; You, S. Machine learning methods for modelling the gasification and pyrolysis of biomass and waste. *Renew. Sustainable Energy Rev.* **2022**, *155*, No. 111902.

(33) Gruber, F.; Grahlert, W.; Wollmann, P.; Kaskel, S. Classification of black plastics waste using fluorescence imaging and machine learning. *Recycling* **2019**, *4*, 40.

(34) Yang, Y.; Zhang, X.; Yin, J.; Yu, X. Rapid and nondestructive on-site classification method for consumer-grade plastics based on partable NIR spectrometer and machine learning. *J. Spectrosc.* **2020**, *2020*, No. 6631234.

(35) Stefas, D.; Gyftokostas, K.; Bellou, E.; Couris, S. Laser-induced breakdown spectroscopy assisted by machine learning for plastics/polymers identification. *Atoms* **2019**, *7*, 79−92.

(36) Lu, H.; Diaz, D. J.; Czarnecki, N. J.; Zhu, C.; Kim, W.; Shroff, R.; Acosta, D. J.; Alexander, B. R.; Cole, H. O.; Zhang, Y.; Lynd, N. A.; Ellington, A. D.; Alper, H. S. Machine learning-aided engineering of hydrolases for PET depolymerization. *Nature* **2022**, *604*, 662−667.

(37) Yel, E.; Tezel, G.; Uymaz, S. A. ANN modelling for predicting the water absorption of composites with waste plastic pyrolysis char fillers. *Data Sci. Appl.* **2018**, *1*, 45−51.

(38) Abnisa, F.; Anuar Sharuddin, S. D.; Bin Zanil, M. F.; Wan Daud, W. M. A.; Indra Mahlia, T. M. The Yield Prediction of Synthetic Fuel Production from Pyrolysis of Plastic Waste by Levenberg−Marquardt Approach in Feedforward Neural Networks Model. *Polymer* **2019**, *11*, 1853.

(39) Mighani, M.; Shahi, A.; Antonioni, G.*Catalytic Pyrolysis of Plastic Waste Products: Time Series Modeling Using Least Square Support Vector Machine and Artificial Neural Network*, 2017.

(40) Okada, S.; Ohzeki, M.; Taguchi, S. Efficient partition of integer optimization problems with one-hot encoding. *Sci. Rep.* **2019**, *9*, 13036.

(41) Aieb, A.; Madani, K.; Scarpa, M.; Bonaccorso, B.; Lefsih, K. A new approach for processing climate missing databases applied to daily rainfall data in Soummam watershed, Algeria. *Heliyon* **2019**, *5*, No. e01247.

(42) Mani, T.; Hauk, A.; Walter, U.; Burkhardt-Holm, P. Microplastics profile along the Rhine River. *Sci. Rep.* **2015**, *5*, 17988.

(43) Vriend, P.; Calcar, C. V.; Kooi, M.; Landman, H.; Pikaar, R.; Emmerik, T. V. Rapid Assessment of Floating Macroplastic Transport in the Rhine. *Front. Mar. Sci.* **2020**, *7*, 10.

(44) Zhang, S. Nearest neighbor selection for iteratively kNN imputation. *J. Syst. Software* **2012**, *85*, 2541−2552.

(45) Nadkarni, P.Core Technologies: Data Mining and "Big Data". In *Clinical Research Computing A Practitioner's Handbook*, 2016; pp 187 −204.

(46) Van Rossum, G.; Drake, F. L.*Python 3 Reference Manual*; CreateSpace: Scotts Valley, CA, 2009.

(47) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, E. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825−2830.

(48) Williams, E. A.; Williams, P. T. The pyrolysis of individual plastics and a plastic mixture in a fixed bed reactor. *J. Chem. Technol. Biotechnol.* **1997**, *70*, 9−20.

(49) Alberg, D.; Last, M.; Kandel, A. Knowledge discovery in data streams with regression tree methods. *WIREs, Data Mini. Knowl. Discovery* **2012**, *2*, 69−78.

(50) MacKay, D. J. C., Hyperparameters: Optimize, or Integrate Out? In *Maximum Entropy and Bayesian Methods: Santa Barbara, California, U.S.A., 1993*, Heidbreder, G. R., Ed.; Springer Netherlands: Dordrecht, 1996; pp 43−59.

(51) Stockwell, D. R.; Peterson, A. T. Effects of sample size on accuracy of species dsitribution models. *Ecol. Model.* **2002**, *148*, 1−13.

(52) Olive, D. J.Multiple Linear Regression. In *Linear Regression*, Olive, D. J., Ed.; Springer International Publishing: Cham, 2017; pp 17−83.

(53) Sakata, Y.; Uddin, M. A.; Muto, A. Degradation of polyethylene and polypropylene into fuel oil by using solid acid and non-acid catalysts. *J. Anal. Appl. Pyrolysis* **1999**, *51*, 135−155.

(54) Zolghadr, A.; Sidhu, N.; Mastalski, I.; Facas, G.; Maduskar, S.; Uppili, S.; Go, T.; Neurock, M.; Dauenhauer, P. J. On the Method of Pulse-Heated Analysis of Solid Reactions (PHASR) for Polyolefin Pyrolysis. *ChemSusChem* **2021**, *14*, 4214−4227.

(55) Scott, D. S.; Czernik, S. R.; Piskorz, J.; Radlein, D. S. A. G. Fast Pyrolysis of Plastic Wastes. *Energy Fuels* **1990**, *4*, 407−411.

(56) Encinar, J. M.; González, J. F. Pyrolysis of synthetic polymers and plastic wastes. Kinetic study. *Fuel Process. Technol.* **2008**, *89*, 678−686.

(57) Odejobi, O. J.; Oladunni, A. A.; Sonibare, J. A.; Abegunrin, I. O. Oil yield optimization from co-pyrolysis of low-density polyethylene (LDPE), polystyrene (PS) and polyethylene terephthalate (PET) using simplex lattice mixture design. *Fuel Commun.* **2020**, *2−5*, No. 100006.

(58) Dhahak, A.; Hild, G.; Rouaud, M.; Mauviel, G.; Burkle-Vitzthum, V. Slow pyrolysis of polyethylene terephthalate: Online monitoring of gas production and quantitative analysis of waxy products. *J. Anal. Appl. Pyrolysis* **2019**, *142*, No. 104664.

(59) Smith, R. L.; Takkellapati, S.; Riegerix, R. C. Recycling of Plastics in the United States: Plastic Material Flows and Polyethylene Terephthalate (PET) Recycling Processes. *ACS Sustainable Chem. Eng.* **2022**, *10*, 2084−2096.

(60) Wang, C.-Q.; Wang, H.; Liu, Y.-N. Separation of polyethylene terephthalate from municipal waste plastics by froth flotation for recycling industry. *Waste Manage.* **2015**, *35*, 42−47.

(61) Belden, E. R.; Kazantzis, N. K.; Reddy, C. M.; Kite-Powell, H.; Timko, M. T.; Italiani, E.; Herschbach, D. R. Thermodynamic feasibility of shipboard conversion of marine plastics to blue diesel for self-powered ocean cleanup. *Proc. Natl. Acad. Sci. U. S. A.* **2021**, *118*, No. e2107250118.

(62) Calcar, C. J. V.; Emmerik, T. H. M. V. Abundance of plastic debris across European and Asian rivers. *Environ. Res. Lett.* **2019**, *14*, 124051.

(63) Wal, M. V. D.; Meulen, M. V. D.; Tweehuijsen, G.; Peterlin, M.; Palatinus, A.; Virsek, M. K.; Coscia, L.; Krzan, A.*SFRA0025: Identification and Assessment of Riverine Input of (Marine) Litter; Final Report for the European Commission DG Environment under Framework Contract No ENV.D.2/FRA/2012/0025*, 2015.

(64) IEA *Production forecast of thermoplastics worldwide from 2020 to 2050, by type (in million metric tons)*. https://www.statista.com/statistics/1192886/thermoplastics-production-volume-by-type-globally/ (accessed August 25, 2022).

(65) *Carbon Dioxide Emissions Coefficients*. https://www.eia.gov/environment/emissions/co2_vol_mass.php (accessed August 2022).

(66) *IEA Global Energy Review*, 2021. https://www.iea.org/reports/global-energy-review-2021 (accessed August 2022).

(67) Papadis, E.; Tsatsaronis, G. Challenges in the decarbonization of the energy sector. *Energy* **2020**, *205*, No. 118025.