

Environmental Chemistry

Predicting Primary Biodegradation of Petroleum Hydrocarbons in Aquatic Systems: Integrating System and Molecular Structure Parameters using a Novel Machine-Learning Framework

Craig Warren Davis,^{a,*} Louise Camenzuli,^b and Aaron D. Redman^a^aExxonMobil Biomedical Sciences, Annandale, New Jersey, USA^bExxonMobil Petroleum & Chemical, Machelen, Belgium

Abstract: Quantitative structure–property relationship (QSPR) models for predicting primary biodegradation of petroleum hydrocarbons have been previously developed. These models use experimental data generated under widely varied conditions, the effects of which are not captured adequately within model formalisms. As a result, they exhibit variable predictive performance and are unable to incorporate the role of study design and test conditions on the assessment of environmental persistence. To address these limitations, a novel machine-learning System-Integrated Model (HC-BioSIM) is presented, which integrates chemical structure and test system variability, leading to improved prediction of primary disappearance time (DT50) values for petroleum hydrocarbons in fresh and marine water. An expanded, highly curated database of 728 experimental DT50 values (181 unique hydrocarbon structures compiled from 13 primary sources) was used to develop and validate a supervised model tree machine-learning model. Using relatively few parameters (6 system and 25 structural parameters), the model demonstrated significant improvement in predictive performance (root mean square error = 0.26, $R^2 = 0.67$) over existing QSPR models. The model also demonstrated improved accuracy of persistence (P) categorization (i.e., “Not P/P/vP”), with an accuracy of 96.8%, and false-positive and -negative categorization rates of 0.4% and 2.7%, respectively. This significant improvement in DT50 prediction, and subsequent persistence categorization, validates the need for models that integrate experimental design and environmental system parameters into biodegradation and persistence assessment. *Environ Toxicol Chem* 2022;41:1359–1369. © 2022 ExxonMobil Biomedical Sciences, Inc. *Environmental Toxicology and Chemistry* published by Wiley Periodicals LLC on behalf of SETAC.

Keywords: Biodegradation; Hydrocarbon; Machine learning; Environmental modeling; Persistent compounds; Quantitative structure–property relationship

INTRODUCTION

Evaluation of chemical degradation processes, particularly biodegradation, is a key element of chemical regulatory management around the globe (Leonards, 2017). Experimental data and predictive models for biodegradation are used extensively in prioritization and evaluation of persistent, bio-accumulative, and toxic chemicals (European Commission, 2008), in regulatory risk evaluation of new and existing

chemicals (Canada, 2005; Davies, 1988; Mo, 2017; US Environmental Protection Agency [USEPA], 2012, 2013), in remediation of contaminated sites (Essaid et al., 2003; Ossai et al., 2020; Ren et al., 2018), and in oil spill response modeling (Atlas, 1995; North et al., 2015; Socolofsky et al., 2019; Spaulding, 2017). Regulatory evaluations and associated persistence criteria typically focus on primary biodegradation rates in environmental media (e.g., water, soil, and sediment; European Chemicals Agency [ECHA], 2017). These values are most often reported as disappearance times (DT50) when kinetics are unknown or biphasic (e.g., presence of a lag phase), or half-lives ($t_{1/2}$) when first-order kinetics are observed. The technical complexity, analytical challenges, and costs associated with environmental biodegradation testing (Shrestha et al., 2020; Whale et al., 2021), particularly the development of reliable primary degradation rates (see Organisation for Economic Co-operation and Development [OECD] test guidelines 307

This article includes online-only Supporting Information.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

* Address correspondence to craig.w.davis@exxonmobil.com

Published online 9 March 2022 in Wiley Online Library

(wileyonlinelibrary.com).

DOI: 10.1002/etc.5328

[2002a], 308 [2002b], and 309 [2004]) in multiple environmental media, have led to the advancement of nontesting methods, including the development of quantitative structure–property relationships (QSPRs) to estimate these properties. Over the last several decades, models have been developed to predict the biodegradation of chemicals under laboratory (e.g., BIOWIN, CATABOL, CATALOGIC; Aronson et al., 2006; Boethling et al., 1994; Dimitrov et al., 2007; Howard et al., 1992, 2005; Jaworska et al., 2002; Meylan et al., 2007) as well as environmental conditions (Pizzo et al., 2016), including the development of a model specifically for petroleum hydrocarbons, BioHCWin (Howard et al., 2005).

The BioHCWin model, developed by Howard et al. (2005), utilizes chemical structural fragments to predict first-order primary biodegradation half-lives (t_h) for petroleum hydrocarbons in aquatic (freshwater and marine) systems. The model was trained ($n = 121$) and validated ($n = 54$) for petroleum hydrocarbon structures encompassing a range of molecular sizes, structural moieties, and physical chemical properties. Experimental degradation data were collected and evaluated from multiple literature sources, representing a diverse set of test conditions (inoculum source, loading rates, temperature, and environmental media). However, in the development and calibration of the model, single recommended t_h values were derived for each hydrocarbon, obfuscating the contributions of test conditions and experimental design on observed variability in biodegradation rates. Consequently, the resulting model predictions are challenging to interpret and cannot be easily compared with experimental data (Prosser et al., 2016).

More recently, machine-learning QSPR models have been developed that predict primary half-lives (t_h) in water, soil, and sediment compartments (Pizzo et al., 2016). Briefly, the models predict t_h values, using a combination of 2D molecular parameters and structural fragment “alerts” (Ferrari et al., 2011; Lombardo et al., 2014; Mauri et al., 2006; Yap, 2011). The data used to train the models (Benfenati et al., 2019; Gouin et al., 2004) consist of a mixture of categorical data (transformed to a single representative t_h value) that do not include test system information, as well as measured t_h data with limited information on test system conditions. As such, these models do not incorporate test system and environmental variability, and so fail to address the limitations and uncertainties of the previous generation of QSPR models. Furthermore, although these models offer quantitative predictions for a wide range of chemicals, their algorithms are opaque and complicated, with parameters and workflows that are difficult to interpret and communicate (Ferrari et al., 2011; Lombardo et al., 2014; Mauri et al., 2006; Yap, 2011). Although these considerations do not necessarily impact the utility of a model, ambiguity and lack of transparency in model algorithm, parameters, and mechanism are commonly cited as barriers to regulatory acceptance (OECD, 2019). As a result of these limitations, as well as unreliable performance outside their training sets (Supporting Information, Figures A1 and A2), machine-learning models have not gained widespread regulatory acceptance as an alternative to costly and challenging environmental testing.

For petroleum hydrocarbons, multiple experimental DT50 values are often available, with considerable variability observed between studies. For example, reported DT50 values for pyrene in seawater range from 7 to 257 days (McFarlin et al., 2018; Ribicic, McFarlin, et al., 2018), with a geometric mean value of 24.4 days ($n = 13$, see Excel File in the Supporting Information). This variability, in part, can be attributed to differences in study conditions, including nutrient availability (Breedveld & Sparrevik, 2000; Delille et al., 1998; McFarlin et al., 2014; Pelletier et al., 2004; Santas et al., 1999; Xu & Obbard, 2003, 2004), test substance loading and source (Birch et al., 2017; Hammershøj et al., 2019; Huesemann et al., 2004; Ren et al., 2018), use of solubilizing agents (Brakstad, Ribicic, et al., 2018), temperature (Ribicic, McFarlin, et al., 2018), and system chemistry (Mormile et al., 1994). This presents a significant challenge in understanding and applying a single QSPR prediction within a regulatory context, because the conditions for which the predictions are relevant (i.e., high/low loading, temperature, and even test system—freshwater vs. seawater) are often not well-defined. As a result, a critical feature of future QSPR models is the ability to quantify and contextualize the relative importance of test system design and environmental conditions on biodegradation rates.

With this in mind, the aim of the present study was threefold: (1) to introduce a novel model (machine-learning System-Integrated Model [HC-BioSIM]) for predicting the primary DT50 values of petroleum hydrocarbons that integrates chemical structure and test system parameters, using an expanded database of high-quality, curated, biodegradation data; (2) to compare the HC-BioSIM model with existing biodegradation models; and (3) to identify and understand the key system and chemical structure parameters that contribute to variability in hydrocarbon primary biodegradation studies for use in regulatory assessment and decision-making.

MATERIALS AND METHODS

To construct and systematically evaluate the performance and utility of the HC-BioSIM model, the following workflow was developed. First, freshwater and marine primary biodegradation data were collected from the peer-reviewed literature. Screening criteria were proposed to evaluate the available data, which included consideration of inoculum selection and test design, robustness of study characterization, data analysis methods, and overall relevance for environmental persistence assessment. The curated database was then randomly split into training (80%) and validation (20%) sets for training the HC-BioSIM model as well as benchmarking against existing and alternative modeling approaches (discussed in the *Model performance and interpretation* section). A k -fold cross-validation of the models was conducted to identify potential training set bias and evaluate the stability and generalizability of the models. Performance of the HC-BioSIM model was compared with that of the existing BioHCWin model as well as two alternative models—a system-integrated BioHCWin (SI-BioHCWin) and a biodegradation polyparameter linear free

energy relationships (bio-pp-LFER) model. Comparing the HC-BioSIM model with existing and alternative models using a consistent independent validation dataset (with cross-validation) is critical to systematically assess predictive performance among model architectures of varying complexity.

Primary biodegradation data

Primary hydrocarbon biodegradation data were obtained from the academic literature and published reports (Birch et al., 2018; Brakstad, Ribicic, et al., 2018; Comber et al., 2012; McFarlin et al., 2018; Prince et al., 2007, 2008, 2013, 2016, 2017; Prosser et al., 2016; Ribicic, McFarlin, et al., 2018; Ribicic, Netzer, et al., 2018). Screening criteria were adopted, with minor modifications, from Brown et al. (2020). Criteria were selected to allow for comprehensive characterization of the experimental design and test system conditions needed for model parameterization as well as to clearly define the environmental applicability domain of the model. A detailed description of the screening criteria can be found in Section A.2 of the Supporting Information. A summary of studies that meet the indicated screening criteria is presented in Table 1.

The applied screening criteria (and their associated test system parameters) have been previously identified as relevant factors that contribute to the degradation rate of hydrocarbons in environmental systems (Birch et al., 2018; Brown et al., 2020; Wang et al., 2018). These parameters are commonly reported and are available in most published studies. Furthermore, these include continuous quantitative and discrete categorical parameters that can be readily utilized by machine-learning algorithms.

The final curated database consisted of 728 DT50 values for 181 unique petroleum hydrocarbons. Studies ranged in incubation temperature from 2 to 21°C and included freshwater, seawater, and activated sludge inoculum. Multiple dosing methods, as well as a wide range of hydrocarbon source materials, were included (Table 1). The complete experimental database was compiled, together with study design and test system details, in an Excel File in the Supporting Information. At present, equivalent experimental databases and screening criteria are not available for petroleum hydrocarbons in soil and sediment systems. Extension of the model to soil and sediment systems, including the curation of analogous high-quality DT50 databases for petroleum hydrocarbons, will be addressed in a companion manuscript.

Model development

HC-BioSIM. Several supervised machine-learning algorithms, including *k*-nearest neighbor, Naïve Bayesian, random forest, and model tree methods were considered for model development. A model framework was desired that balanced improved accuracy of prediction with the need for transparency in computational methodology, applicability domain, and mechanistic interpretation of the model and its output (Gramatica, 2007; OECD, 2019). As a result, algorithms that leverage boosting (e.g., random forest) or hidden layers (e.g., neural networks) were screened out, along with classification or categorical algorithms (e.g., Naïve-Bayes). Decision tree algorithms offer improved flexibility in model structure over traditional modeling approaches (e.g., BioHCWin), while remaining highly interpretable and easy to communicate. Ultimately, the Cubist model tree algorithm (Kuhn et al., 2012),

TABLE 1: Summary of studies that met screening criteria, including relevant study information used in model development and number of data points (No.)

Test media/inoculum source	Hydrocarbon source	Test temperature (°C)	Dosing method	Use of dispersant	No.	References
Freshwater	Gasoline	21	Direct	N	110	Prince et al. (2007)
	B20 diesel	21	Direct	N	72	Prince et al. (2008)
	Defined mixture	20	P.D.	N	33, 22	Prosser et al. (2016), Birch et al. (2018)
Seawater	Crude oil ^a	2	Direct	N	44	McFarlin et al. (2018)
		5	Direct	Y	80, 14, 32	Brakstad, Ribicic, et al. (2018), Prince et al. (2016), Ribicic et al. (2018)
		5–13	Direct	Y	107	Ribicic et al. (2018)
		8	Direct	Y	24	Prince et al. (2013)
		21	Direct	Y	69	Prince et al. (2017)
	Defined mixture	20	P.D.	N	29, 25	Prosser et al. (2016), Birch et al. (2018)
		20	P.D.	N	18	Comber et al. (2012)
Activated sludge	Produced water ^b	13	P.D.	N	10	Lofthus et al. (2018)
	Defined mixture	20	P.D.	N	39	Birch et al. (2018)
	Total: 728					

^aSeveral crude oil datasets were compiled; a complete characterization of the test substances and the experimental designs is available in the Excel File in the Supporting Information.

^bTest material: Produced water containing oil droplets and oil-coated particulates collected from offshore drilling operations in the North Sea. Complete documentation of screening criteria, study designs, test system parameters, and additional notes are provided in Section A2 of the Supporting Information. P.D. = passively dosed.

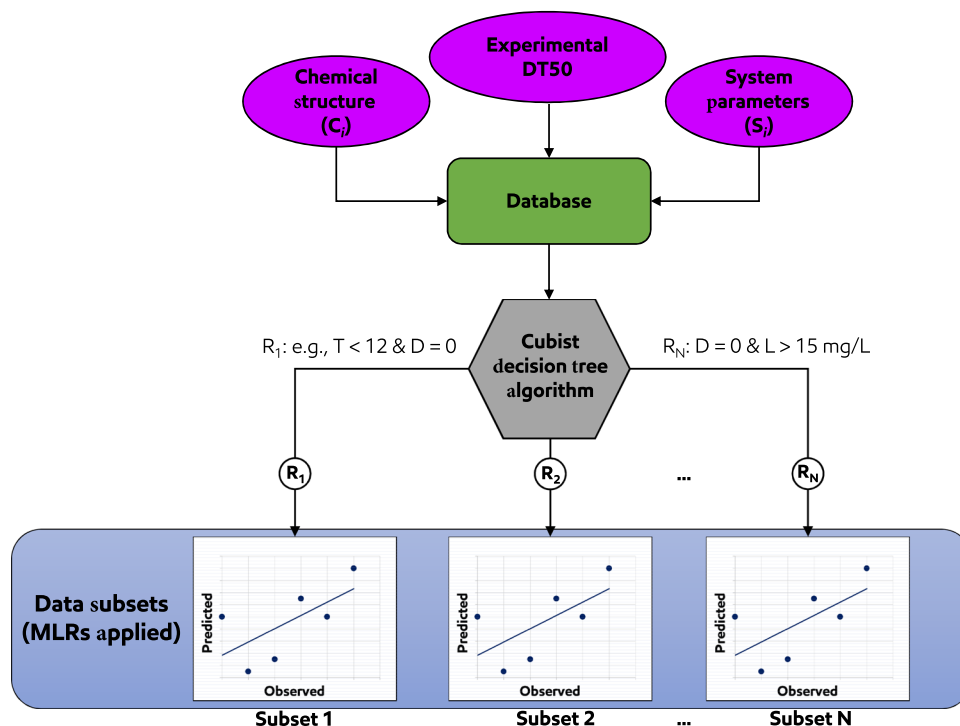


FIGURE 1: Schematic diagram of the System-Integrated Model (HC-BioSIM) cubist decision tree machine-learning workflow. User-defined input includes experimental disappearance time (DT50) values (labels used to train the model), chemical structure (C_i), and system (S_i) parameters. Model-defined rules (R_i) for parsing the dataset are indicated by white circles, and the blue box indicates terminal subset “nodes,” where multiple linear regressions (MLRs) are applied, resulting in a prediction of DT50 values for that subset. Example rules are included for illustrative purpose.

an extension of the M5 algorithm developed by Quinlan (Quinlan, 1993), was selected.

The Cubist model tree develops a single set of “rules” that parse training data into subsets using a set of provided chemical and/or system parameters. The number of rule-based subsets, as well as the relevant parameters, are selected by the algorithm to minimize the entropy of the entire system. This is accomplished by maximizing the similarity between data points within each subset. Redundant rules are trimmed or combined by the model in a postdevelopment “pruning” step to minimize the complexity of the final decision tree model. A unique multiple linear regression (MLR) is then applied to each subset to predict the DT50 values. A critical element of the Cubist algorithm is that both the classification scheme (the rules) and the resulting MLRs are easily communicated and computationally transparent. A generalized workflow for the Cubist decision tree model development is illustrated in Figure 1.

Recently, a set of parameters (ToxPrint) for describing molecular structures of both organic and inorganic chemistries has been developed (Yang et al., 2015) and made freely available by the USEPA within the CompTox Dashboard platform (Williams et al., 2017). These parameters have been used as a basis for read-across assessment, structural similarity calculations, and predictions of various in vitro toxicological endpoints (Drwal et al., 2015; Hur et al., 2017; Helman, Shah, & Patlewicz, 2019; Helman, Shah, Williams, et al., 2019). The ToxPrint parameters were selected for use in model parametrization because they are clearly defined, easily interpreted, and cover a

broad range of chemical structures. For each unique chemical, a “fingerprint” can be constructed as a vector of binary values (0 or 1) indicating the absence or presence of particular structural fragments. These fingerprints can then be combined with additional relevant parameters (i.e., test system parameters) for training the HC-BioSIM model.

The compiled petroleum hydrocarbon DT50 database included 55 unique ToxPrint chemical structural fragment parameters. These parameters were further manually curated to remove redundant fragments, reducing the total number to 45. A detailed description of this curation is available in Section A3 of the Supporting Information. Test system parameters were selected on the basis that they were readily available, frequently reported in literature and regulatory study reports, and are observed to influence observed DT50 values in environmental systems. The selected test system parameters included the test temperature (T ; °C), the hydrocarbon loading rate (L ; mg/L), the hydrocarbon source material (K ; represented by the kinematic viscosity of the source material [cSt] and described in Section A4 of the Supporting Information), the dispersant treatment rate (D ; kg dispersant/kg test substance), and the inoculum source (m_j , where j = freshwater, marine water, or activated sludge).

Chemical structural fragments and test system parameters (D , K , T , L , and m_j) were combined to create a pool of 52 potential parameters to be used in the development of the model rules and MLRs, discussed previously in this section. Model optimization, cross-validation, and relevant statistical analyses were performed using the cubist() package in

R (Ver 3.6.1), unless otherwise stated. A detailed summary of the HC-BioSIM model output is presented in Section A5 of the Supporting Information.

Evaluation of alternative QSPR models (SI-BioHCWin and bio-pp-LFER). Whereas the HC-BioSIM model represents a significant step-change improvement in predicting the environmental biodegradation of petroleum hydrocarbons, it also represents a significant increase in model complexity. Thus it is prudent that we evaluate the added value of such a model framework against more traditional modeling approaches. To this end, two alternative linear models were evaluated to validate the need for, and ultimate selection of, the HC-BioSIM model moving forward.

Briefly, the SI-BioHCWin model modifies the existing BioHCWin (t_r) prediction for a given hydrocarbon structure using a simple linear additive framework (see Supporting Information, Equation SI-1) to incorporate the contributions from each of the system parameters (described previously in the *Primary degradation data* section). Coefficients for test system model parameters are estimated via MLR. This model represents a mechanistically simple and direct approach to the incorporation of system parameters, with relatively few estimated parameters (seven or less). A complete description of the methodology can be found in Section A4 of the Supporting Information.

An alternative approach is to directly estimate the contributions of chemical structure and system parameters simultaneously. Briefly, the bio-pp-LFER model utilizes molecular parameters previously described by Abraham and Acree (2010) as well as computed HOMO-LUMO energies (as a surrogate for potential chemical reactivity in biological systems (Abraham & Acree, 2010; Mekenyan & Veith, 1994; Siraki et al., 2005; Veith et al., 1995; Voutchkova et al., 2011) to describe the chemical contributions to the observed DT50 values. System parameters are described using the same additive framework as described for the SI-BioHCwin model just described (see Equation SI-4 in the Supporting Information). In this case, chemical and system coefficients are estimated simultaneously via MLR. This approach is similar to those previously applied to environmental partitioning systems (Endo & Goss, 2014) as well as more recently for biotransformation processes (Kuo & Di Toro, 2013) of neutral organic chemicals. This approach introduces additional estimated parameters (12 or less), but is similarly simple with respect to the use of a single MLR model. A complete description of the methodology can be found in Section A4 of the Supporting Information.

Model calibration and data analysis. All three models were trained and validated using the expanded biodegradation DT50 database. The data were split randomly into training (80%) and validation (20%) sets. A random seed was utilized to ensure identical training/validation sets for the three models as well as reproducibility of the results from the R code (see Section A.6 of the Supporting Information) This was done to ensure identical representation of limited datasets (e.g., activated sludge DT50 data) and minimize any potential bias in the training and validation sets when comparing the models. All

model development, calibration, analysis, and visualization was performed in R Ver 3.6.1.

Model performance was evaluated using the root mean square error (RMSE) as well as the Pearson correlation coefficient (R^2). The RMSE values were computed as follows:

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (\log(\text{DT50}_{\text{pred}}) - \log(\text{DT50}_{\text{exp}}))^2} \quad (1)$$

where N is the total number of observations, $\text{DT50}_{\text{pred}}$ is the predicted DT50 (days), and DT50_{exp} is the experimental DT50 (days). A logarithmic transformation was applied to ensure equal weighting of predictive errors across the large range of DT50 values as well as to provide additional clarity with respect to communicating uncertainties in the predicted values. For example, logarithmic RMSE (Equation 1) values of 0.3 and 0.5 represent two- and threefold average errors in the predicted DT50, respectively.

Finally, to assess the generalizability of the models and to evaluate any potential training set bias, a k -fold cross-validation ($k=5$) was performed on all three models. Mean and standard deviation of RMSE and R^2 values for the five validation folds were used to compare the different models. Mean and standard deviations of the rule and end-node parameter usage (%; HC-BioSIM model) and estimated coefficients (SI-BioHCWin and bio-pp-LFER models) were computed to assess parameter importance. A complete description of the cross-validation methodology and results is presented in Section A7 of the Supporting Information.

RESULTS AND DISCUSSION

Model performance and interpretation

Performance of the HC-BioSIM model was benchmarked against BioHCWin as well as the alternative models for both the training and validation datasets. Figure 2 shows a comparison of predicted and experimental DT50 (day) values for the HC-BioSIM and the original BioHCWin model, using the expanded aquatic biodegradation database. Data were separated and visualized by hydrocarbon class, to identify potential outliers or systematic bias that might impact the overall model performance. Model statistics including RMSE, R^2 , and cross-validation results for the HC-BioSIM and BioHCWin models are summarized in Table 2. Model results and associated discussions of the SI-BioHCWin and the bio-pp-LFER model are included in Section A4 of the Supporting Information.

The ability of the BioHCWin model to reproduce the experimental DT50 data for the varied test conditions represented in the expanded database was poor (Figure 2A). Significant overprediction of DT50 values for naphthenic polycyclic aromatic hydrocarbons (NPAH), polycyclic aromatic hydrocarbons (PAH), and naphthenic diaromatics (NDAH) compounds were observed, with a number of significant outliers (i.e., overpredicted by more than 100-fold). These results are consistent with a previous evaluation of BioHCWin by Prosser et al. (2016). The HC-BioSIM model significantly outperformed the BioHCWin model (Table 2 and Figure 2B), as well as both the SI-BioHCWin and the

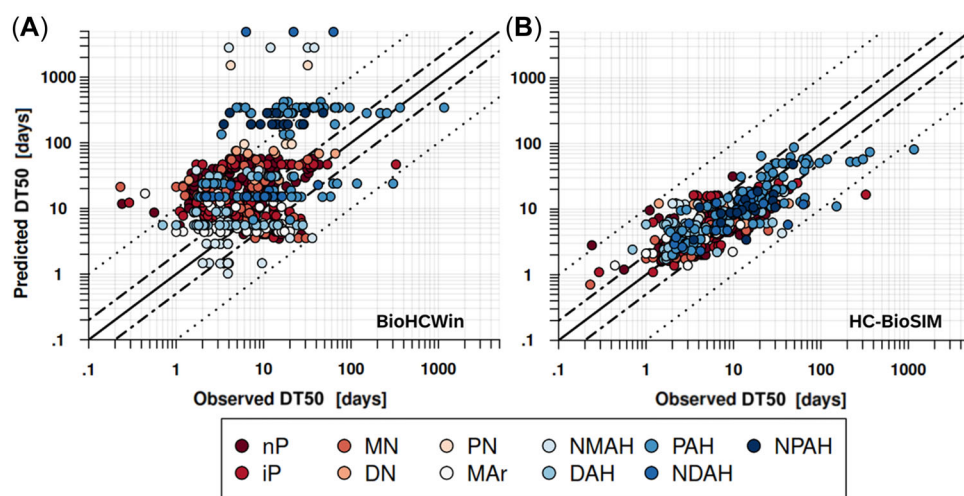


FIGURE 2: Predicted versus observed experimental disappearance time (DT50; in days) for the (A) BioHCWin and (B) System-Integrated Model (HC-BioSIM) models. Solid line represents 1:1 agreement, and semidashed and dashed lines represent 3x and 10x errors in predictions, respectively. Colors correspond to hydrocarbon classes: *n*-paraffins (nP), iso-paraffins (iP), mono-naphthenics (MN), di-naphthenics (DN), poly-naphthenics (PN), mono-aromatics (MAr), naphthenic mono-aromatics (NMAH), di-aromatics (DAH), polyaromatics (PAH), naphthenic di-aromatic (NDAH), and naphthenic polyaromatics (NPAH).

bio-pp-LFER models (Supporting Information, Figure A3) for both the training and validation datasets. The average predicted DT50 error for the HC-BioSIM model (~1.8x) represents a 3-fold improvement over that of the BioHCWin model (~5.6x). Furthermore, bias as a function of hydrocarbon class or carbon number was not observed (Figure 3E–F). There were far fewer significant outliers for the HC-BioSIM model (four DT50 predictions with errors greater than 10x). These four outliers represent a wide range of study conditions, further supporting a lack of systematic bias in the model (Figure 3A–D).

Finally, it is critical to note that despite poorer performance relative to the HC-BioSIM model, the SI-BioHCWin and bio-pp-LFER models showed improvement over the BioHCWin model. This finding supports the general principle that predictive performance and model utility can be improved if detailed system information and curation of data are integrated into model development. This can be leveraged to provide insight and identify opportunities for model improvement for a broader range of chemistries and processes in which reliable

experimental data and associated system information are available.

System parameters and residual analyses

System parameters. Although the results of the model tree algorithm are more complicated to visually interpret, parameter importance and mechanistic trends can be observed from an analysis of the rule structure and parameter usage within the submodels. Model tree algorithms are an example of a “greedy” algorithm (Cormen et al., 2009)—they sequentially separate data using the most impactful parameter, maximizing information gain and minimizing system entropy. Consequently, the use of parameters in constructing model “rules,” as well as their presence in the end-node MLRs, is correlated to their relative importance. This allows some mechanistic insights to be drawn. The model rules and summary statistics for the end-node MLRs are summarized in Table 3. A list of model parameters and their usage (%), as well as the complete end-node MLRs, is presented in Section A5 of Supporting Information.

The HC-BioSIM model utilizes four of the seven available system-specific parameters (*D*, *L*, *T*, *K*) as well as a single structural parameter (a nonlinear three-ring PAH fragment, PAH_{NL}) to develop the rules to subset the experimental DT50 training database. Furthermore, six of the seven available system parameters and 25 of the 45 structural parameters are used in the MLRs across the subsets to predict DT50 values. Although the number of model parameters (31) is significantly larger than the SI-BioHCWin (6) and bio-pp-LFER (10) models, it is comparable to the number of fragment values utilized within the BioHCWin model (32).

It is interesting to note that the average model errors (E) for the individual data subsets (Table 3) are comparable, with the exception of subsets #4 and #9. These subsets are unique in that they contain all passively dosed data. Larger predictive

TABLE 2: Comparison of HC-BioSIM and BioHCWin model performance for training and validation sets (including *k*-fold cross-validation)

Dataset	No.	HC-BioSIM		BioHCWin	
		RMSE	R^2	RMSE	R^2
Training	582	0.23	0.71	0.76	0.16
Validation	146	0.34	0.52	0.72	0.18
All	728	0.26	0.67	0.75	0.17
CV test fold ^a	146	0.30 ± 0.01 (3.2%)	0.51 ± 0.08 (16%)	0.75 ± 0.05 (6.0%)	0.16 ± 0.03 (18%)

^aMean ± standard deviation (SD) RMSE and R^2 values for the individual test folds (*k* = 5). Coefficients of variation (%) are included in parentheses. RMSE and R^2 values are reported for both models. A complete description of the cross-validation technique, individual fold statistics, and parameters is presented in Section A7 of the Supporting Information. CV = cross validation; RMSE = root mean square error.

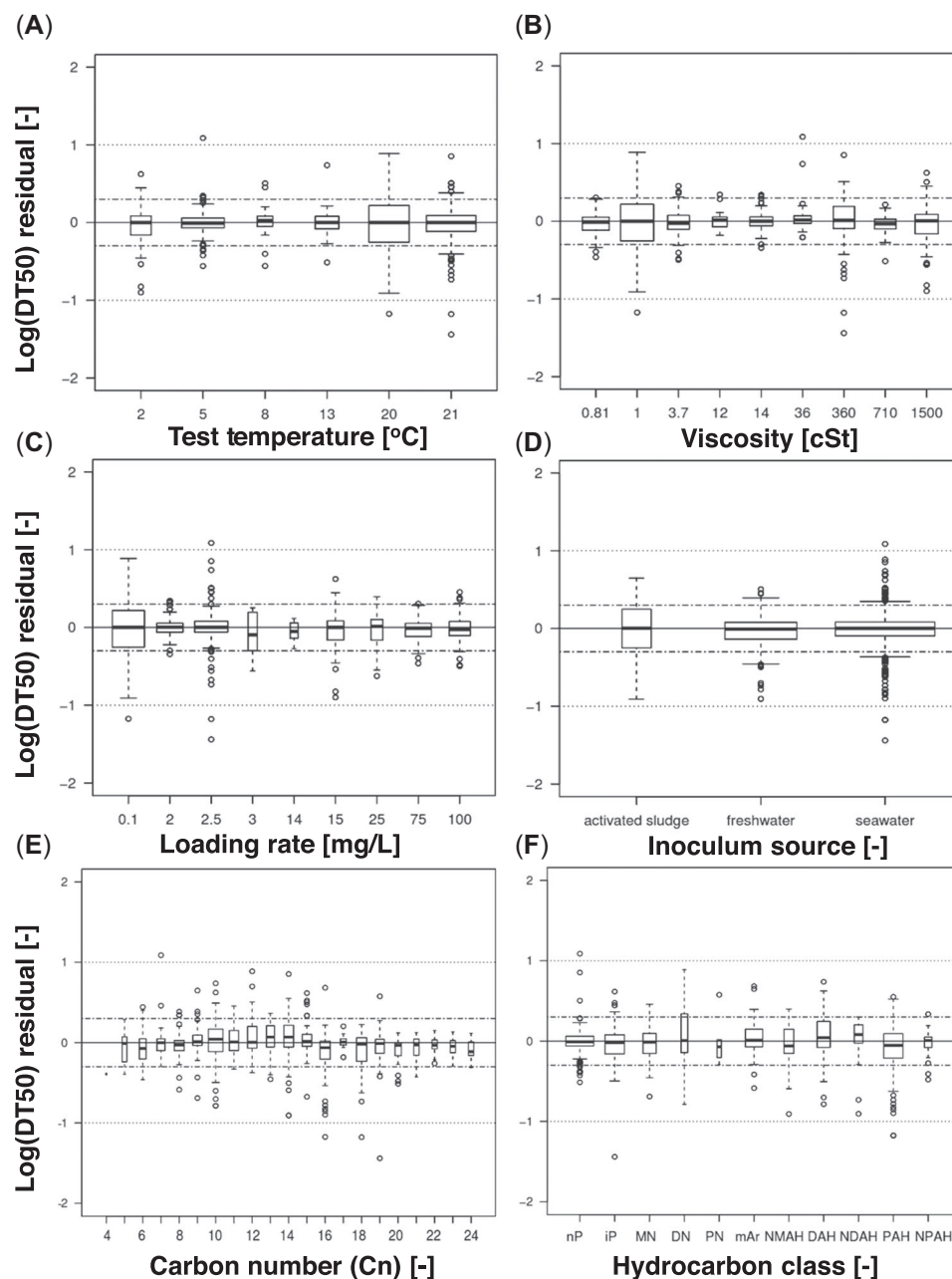


FIGURE 3: Boxplots of logarithmic model residuals (predicted—experimental $\log(\text{DT50})$) as a function of test system parameters (A–D), carbon number (E), and hydrocarbon class (F). Semidashed lines represent a 2-fold predicted error (0.3 log units), and dashed lines represent a 10-fold predicted error (1.0 log unit). Box widths are proportional to the square root of the number of observations. For abbreviations, see Figure 2 legend.

errors for these subsets may be attributed to uncertainty in the appropriate parameterization for passively dosed studies. Although all studies within these subsets are nondispersed ($D = 0$), lower loadings in the passively dosed studies may result in significantly different biodegradation behavior compared with studies conducted at higher loadings (e.g., crude oil at 2–15 mg/L; Hammershøj et al., 2019, 2020). Further investigation is needed to evaluate whether additional system parameters (i.e., a specific system parameter for passively dosed studies) would improve model predictions for these studies. At present, passively dosed DT50 data are limited, and additional

datasets may be required to adequately incorporate any additional parameterization.

The model also identified high hydrocarbon loading rates, regardless of other system parameters, as a unique subset (#2: $L > 15$ mg/L). Previous studies have demonstrated slower overall removal of hydrocarbons at high source loading (Brakstad, Davies, et al., 2018), as well as trends within and between hydrocarbon classes that differ with hydrocarbon loading (i.e., 75 mg/L direct loading of gasoline vs. low concentrations of passively dosed gasoline-range hydrocarbons; Prince et al., 2007; Prosser et al., 2016). This observation

TABLE 3: Summary of HC-BioSIM model subsets (S), rules (R), number of data points (No.), logarithmic average prediction error (E), and brief descriptions of data subsets

Subset (S)	Rules (R)	No.	Average predictive error (E)	Description of data subset
1	$D > 0$ $L \leq 15$ $T > 8$ $PAH_{NL} = 0$	51	0.12	Dispersed, low loading, mid-high temperature, no PAH_{NL}
2	$L > 15$	162	0.14	High loading
3	$D > 0$ $T > 8$ $PAH_{NL} = 1$	16	0.17	Dispersed, mid-high temperature, PAH_{NL}
4	$D = 0$ $L \leq 15$ $T > 13$ $PAH_{NL} = 0$	129	0.37	Dispersed, low loading, high temperature, no PAH_{NL}
5	$K \leq 14.4$ $T \leq 8$ $PAH_{NL} = 0$	78	0.10	Low-viscosity HC substrate, low-temperature, no PAH_{NL}
6	$D = 0$ $T \leq 13$ $PAH_{NL} = 0$	54	0.16	Nondispersed, mid-low temperature, no PAH_{NL}
7	$K > 14.4$ $D > 0$ $T \leq 8$ $PAH_{NL} = 0$	51	0.09	High-viscosity HC substrate, dispersed, low temperature, no PAH_{NL}
8	$T \leq 8$ $PAH_{NL} = 1$	36	0.18	Low temperature, PAH_{NL}
9	$D = 0$ $L \leq 15$ $T > 8$ $PAH_{NL} = 1$	24	0.38	Nondispersed, low loading, mid-high temperature, PAH_{NL}

PAH_{NL} = presence or absence of non-linear 3-ring PAH structural fragment.

supports the selection of hydrocarbon loading as an effective system parameter for describing aquatic DT50 test systems.

Only one structural parameter was identified as significant to rule development for the model. The three-ring PAH_{NL} was used by the model to separate a significant segment of PAH hydrocarbons from the rest of the aromatic and aliphatic hydrocarbon space. This result is not immediately intuitive; however, model predictive performance for PAHs (and in particular, those not containing this structural fragment, e.g., anthracene) does not appear to be impacted by the use of this parameter in generating the model rules. It is possible that the limited aqueous solubility of PAHs necessitates the separation from smaller aromatics, with the more rapid biodegradability of PAHs further separating these structures from their equivalent naphthenic analogs (e.g., polynaphthenics [PN]).

It is important to note that model rule cutoff values (i.e., $L > 15$ mg/L) should be interpreted cautiously, because complete characterization and range of the system variables are likely incomplete. As such, true “criteria” for these rules may lie somewhere in between values at which data have previously been obtained. In addition, potential nonlinear interactions between

system parameters (e.g., low temperature and high loading) are not currently represented within the model framework.

Residual analysis. Residual errors of HC-BioSIM predicted DT50 values as a function of test system parameters, hydrocarbon class, and carbon number are presented in Figure 3. Similar analyses were performed for the BioHCWin, SI-BioHCWin, and bio-pp-LFER models and can be found in Section A4 of the Supporting Information, Figures A4–A6.

No systematic bias was observed in the predicted DT50 values as a function of test system parameters, hydrocarbon classes, or carbon number (Figure 3). In a similar analysis by Prosser et al. (2016) the authors observed clear systematic bias in BioHCWin t_h predictions as a function of carbon number, with significant overpredictions at higher carbon numbers across multiple classes, and the most pronounced bias in the normal and iso-paraffin classes. The authors also reported significant systematic predictive bias across several hydrocarbon classes, specifically naphthenic hydrocarbons (mononaphthenics (MN), dinaphthenics (DN), and PN), for which experimental DT50 data are limited.

Categorization of P and vP substances

The ability of the HC-BioSIM model to correctly predict the persistence of hydrocarbons in freshwater and marine systems was evaluated (based on DT50 criteria proposed by the ECHA, 2017), similar to a previous assessment by Prosser et al. (2016) for the BioHCWin model. Outcomes were grouped broadly into three categories: (1) correct predictions, with identical categorization of observed and model-predicted persistence,

TABLE 4: Prediction matrix of persistence categorization based on European Chemicals Agency freshwater and marine compartmental half-life criteria

System	Prediction	Model	
		BioHCWin (%)	HC-BioSIM (%)
Freshwater ^a ($n = 237$)	FN (type II)	0.4	1.3
	Correct	93.7	97.9
	FP (type I)	5.9	0.8
Seawater ^b ($n = 452$)	FN (type II)	1.1	3.3
	Correct	87.6	96.2
	FP (type I)	11.3	0.4
Total ^c ($n = 689$)	FN (type II)	0.9	2.6
	Correct	89.7	96.8
	FP (type I)	9.43	0.6

^aFor freshwater, the European Union Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH) P and vP criteria of 40 and 60 days are used, respectively.

^bFor marine water, the European Union REACH singular P and vP criteria of 60 days are used.

^cActivated sludge primary disappearance time (DT50) values ($n = 39$) were excluded from this evaluation, because their applicability in comparing against either freshwater or marine DT50 criteria is not clear.

Prediction matrices for the SI-BioHCWin and bio-pp-LFER models are presented in Section A8 of the Supporting Information. FN = false negative; FP = false positive.

(2) underprediction of persistence (false negatives), and (3) overprediction of persistence (false positives). A comparison of the HC-BioSIM and BioHCWin categorization results are summarized in Table 4. Results for the SI-BioHCWin and bio-pp-LFER models are presented in Section A8 of the Supporting Information.

The HC-BioSIM model demonstrated improved accuracy of persistence categorization, with an overall correct identification rate of 96.8% compared with 89.7% for the BioHCWin model. False-positive predictions were significantly reduced with the HC-BioSIM model (0.6% vs. 9.4%). However, a slight increase in “false-negative” predictions was observed (2.6% vs. 0.9%). After further investigation, it was seen that half of these data points were obtained from nondispersed, high-viscosity crudes, seven of them at extremely high loading rates ($L = 15$ mg/L). Although these test conditions may be representative for oil spill scenarios, and valuable within a risk assessment context, they should not be considered representative conditions for evaluating the intrinsic persistence properties of a substance. The OECD test guidelines for assessing biodegradation of chemicals in the environment, such as OECD test guideline 309 (2004) for aquatic simulation studies, recommend test concentrations of 0.001–0.1 mg/L. Further exclusion or qualification of these studies resulted in a revised overall false-negative categorization rate of approximately 1.6%. It should be noted that for many of these hydrocarbons, additional DT50 data, obtained under more environmentally representative system conditions, demonstrate agreement between predicted and observed persistence categorization.

Despite significant improvement in DT50 prediction and an enhanced understanding of system effects on biodegradability, a significant challenge remains in the interpretation of these data and models within existing regulatory frameworks. At present, limited guidance is available for developing a single persistence conclusion using multiple DT50 results, representing varied system conditions (e.g., loading, dispersed vs. nondispersed, passively dosed, etc.). One potential strategy for addressing this challenge is to utilize statistical methods (e.g., Monte-Carlo simulations). These methods can sample a range of relevant system parameters, producing a probability distribution of DT50 values and associated “P” categorizations for a given substance. In addition, guidance on ranges of values for environmental and test conditions that are most relevant for persistence assessment could reduce uncertainty and provide clear direction for persistence assessment of petroleum hydrocarbons in complex environmental systems. These considerations would provide a quantitative framework that could be used to supplement the existing “weight of evidence” approaches (Hughes et al., 2020) that are currently used to evaluate the persistence properties of petroleum hydrocarbons.

CONCLUSIONS

Leveraging a large, highly curated, hydrocarbon DT50 database allowed for system- and substance-specific variability in environmental biodegradation rates to be systematically

integrated within a predictive model for the first time. Test system and environmental parameters were shown to be critical factors in the development of model rules as well as in the prediction of DT50 values. The HC-BioSIM model demonstrated significant improvement in quantitative DT50 prediction, as well as persistence categorization for a wide range of experimental test conditions and hydrocarbon structures. These results further reinforce the need to consider environmental and system conditions when persistence data for regulatory evaluation and risk assessment are compared. Finally, the model presented is transparent and easily communicated, addressing several key challenges to regulatory acceptance. It is expected that this approach may be applied to additional biological and chemical processes (i.e., metabolism, abiotic degradation processes) in various media, leveraging existing databases in which sufficient high-quality test system and chemical information can be identified and curated appropriately.

Supporting Information—The Supporting Information is available on the Wiley Online Library at <https://doi.org/10.1002/etc.5328>.

Acknowledgments—The authors would like to thank J.S. Arey, A. Ireland, R. Prince, and T. Parkerton for their guidance, feedback, and helpful discussions. We would also like to thank the manuscript reviewers and editor for their careful review and constructive feedback.

Disclaimer—The authors declare no competing interest. This manuscript has not been previously published, in whole or in part, and is not under consideration by any other journal. All authors are aware of and accept responsibility for the views presented in this manuscript.

Author Contributions Statement—**Craig Warren Davis**: Conceptualization; Methodology; Formal analyses; Writing—original draft. **Louise Camenzuli**: Writing—original draft, review & editing. **Aaron D. Redman**: Writing—original draft, review & editing.

Data Availability Statement—The complete data presented in this manuscript (figures and tables) as well as the chemical and test system parameters, source code for the QSPR model(s), and cross-validation results are available in full in the Supporting Information. Data, associated metadata, and calculation tools are also available from the corresponding author (craig.w.davis@exxonmobil.com).

REFERENCES

- Abraham, M. H., & Acree, W. E., Jr. (2010). Equations for the transfer of neutral molecules and ionic species from water to organic phases. *The Journal of Organic Chemistry*, 75, 1006–1015.
- Aronson, D., Boethling, R., Howard, P., & Stiteler, W. (2006). Estimating biodegradation half-lives for use in chemical screening. *Chemosphere*, 63, 1953–1960.

- Atlas, R. M. (1995). Petroleum biodegradation and oil spill bioremediation. *Marine Pollution Bulletin*, 31, 178–182.
- Benfenati, E., Roncaglioni, A., Lombardo, A., & Manganaro, A. (2019). Integrating QSAR, read-across, and screening tools: The VEGA HUB platform as an example. In H. Hong (Ed.), *Advances in Computational Toxicology* (pp. 365–381). Springer.
- Birch, H., Hammershøj, R., Comber, M., & Mayer, P. (2017). Biodegradation of hydrocarbon mixtures in surface waters at environmentally relevant levels—Effect of inoculum origin on kinetics and sequence of degradation. *Chemosphere*, 184, 400–407.
- Birch, H., Hammershøj, R., & Mayer, P. (2018). Determining biodegradation kinetics of hydrocarbons at low concentrations: Covering 5 and 9 orders of magnitude of K_{ow} and K_{aw} . *Environmental Science & Technology*, 52, 2143–2151.
- Boethling, R. S., Howard, P. H., Meylan, W., Stiteler, W., Beauman, J., & Tirado, N. (1994). Group contribution method for predicting probability and rate of aerobic biodegradation. *Environmental Science & Technology*, 28, 459–465.
- Brakstad, O. G., Davies, E. J., Ribicic, D., Winkler, A., Brønner, U., & Netzer, R. (2018). Biodegradation of dispersed oil in natural seawaters from Western Greenland and a Norwegian fjord. *Polar Biology*, 41, 2435–2450.
- Brakstad, O. G., Ribicic, D., Winkler, A., & Netzer, R. (2018). Biodegradation of dispersed oil in seawater is not inhibited by a commercial oil spill dispersant. *Marine Pollution Bulletin*, 129, 555–561.
- Breedveld, G. D., & Sparrevik, M. (2000). Nutrient-limited biodegradation of PAH in various soil strata at a creosote contaminated site. *Biodegradation*, 11, 391–399.
- Brown, D. M., Camenzuli, L., Redman, A. D., Hughes, C., Wang, N., Vaipoulou, E., Saunders, D., Villalobos, A., & Lington, S. (2020). Is the Arrhenius-correction of biodegradation rates, as recommended through REACH guidance, fit for environmentally relevant conditions? An example from petroleum biodegradation in environmental systems. *Science of the Total Environment*, 732, 139293. <https://doi.org/10.1016/j.scitotenv.2020.139293>
- Canada. (2005). New substances notification regulations (chemicals and polymers). SOR/2005–247. In Government of Canada (Ed.), *Canada Gazette Part 2. Vol 139* (pp. 1864–1903).
- Comber, M., Haan, K. D., Djemel, N., & Eadsforth, C. V. (2012). Primary biodegradation of petroleum hydrocarbons in seawater. *Concawe Report*.
- Cormen, T. H., Leiserson, C. E., Rivest, R. L., & Stein, C. (2009). *Introduction to algorithms*. MIT Press.
- Davies, J. (1988). CEPA—The Canadian Environmental Protection Act. *Journal of the Air Pollution Control Association*, 38, 1111–1113. <https://doi.org/10.1080/08940630.1988.10466452>
- Delille, D., Bassères, A., & Dessommes, A. (1998). Effectiveness of bioremediation for oil-polluted Antarctic seawater. *Polar Biology*, 19, 237–241.
- Dimitrov, S., Pavlov, T., Nedelcheva, D., Reuschenbach, P., Silvani, M., Bias, R., Comber, M., Low, L., Lee, C., Parkerton, T., & Mekenyan, O. (2007). A kinetic model for predicting biodegradation. *SAR and QSAR in Environmental Research*, 18, 443–457.
- Drwal, M. N., Siramshetty, V. B., Banerjee, P., Goede, A., Preissner, R., & Dunkel, M. (2015). Molecular similarity-based predictions of the Tox21 screening outcome. *Frontiers in Environmental Science*, 3, 54.
- European Chemicals Agency. (2017). Guidance on information requirements and chemical safety assessment (Chapter R.11: PBT/vPvB assessment). <https://doi.org/10.2823/128621>
- Endo, S., & Goss, K.-U. (2014). Applications of polyparameter linear free energy relationships in environmental chemistry. *Environmental Science & Technology*, 48, 12477–12491.
- Essaid, H. I., Cozzarelli, I. M., Eganhouse, R. P., Herkelrath, W. N., Bekins, B. A., & Delin, G. N. (2003). Inverse modeling of BTEX dissolution and biodegradation at the Bemidji, MN crude-oil spill site. *Journal of Contaminant Hydrology*, 67, 269–299.
- European Commission. (2008). Commission Regulation (EC) No 440/2008 of 30 May 2008 laying down test methods pursuant to Regulation (EC) No 1907/2006 of the European Parliament and of the Council on the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH).
- Ferrari, T., Giuseppina, G., Nazanin, G. B., & Emilio, B. (2011). Mining toxicity structural alerts from SMILES: A new way to derive Structure Activity Relationships. *2011 IEEE Symposium on Computational Intelligence and Data Mining (CIDM)* (pp. 120–127).
- Gouin, T., Cousins, I., & Mackay, D. (2004). Comparison of two methods for obtaining degradation half-lives. *Chemosphere*, 56, 531–535.
- Gramatica, P. (2007). Principles of QSAR models validation: Internal and external. *QSAR & Combinatorial Science*, 26, 694–701.
- Hammershøj, R., Birch, H., Redman, A. D., & Mayer, P. (2019). Mixture effects on biodegradation kinetics of hydrocarbons in surface water: Increasing concentrations inhibited degradation whereas multiple substrates did not. *Environmental Science & Technology*, 53, 3087–3094.
- Hammershøj, R., Sjøholm, K. K., Birch, H., Brandt, K. K., & Mayer, P. (2020). Biodegradation kinetics testing of two hydrophobic UVCBs—Potential for substrate toxicity supports testing at low concentrations. *Environmental Science: Processes & Impacts*, 22, 2172–2180. <https://doi.org/10.1039/D0EM00288G>
- Helman, G., Shah, I., & Patlewicz, G. (2019). Transitioning the generalised read-across approach (GenRA) to quantitative predictions: A case study using acute oral toxicity data. *Computational Toxicology*, 12, 100097.
- Helman, G., Shah, I., Williams, A. J., Edwards, J., Dunne, J., & Patlewicz, G. (2019). Generalized read-across (GenRA): A workflow implemented into the EPA CompTox chemicals dashboard. *ALTEX-Alternatives to Animal Experimentation*, 36, 462–465.
- Howard, P., Meylan, W., Aronson, D., Stiteler, W., Tunkel, J., Comber, M., & Parkerton, T. F. (2005). A new biodegradation prediction model specific to petroleum hydrocarbons. *Environmental Toxicology and Chemistry: An International Journal*, 24, 1847–1860.
- Howard, P. H., Stiteler, W. M., Meylan, W. M., Hueber, A. E., Beauman, J. A., Larosche, M. E., & Boethling, R. S. (1992). Predictive model for aerobic biodegradability developed from a file of evaluated biodegradation data. *Environmental Toxicology and Chemistry: An International Journal*, 11, 593–603.
- Huesemann, M. H., Hausmann, T. S., & Fortman, T. J. (2004). Does bio-availability limit biodegradation? A comparison of hydrocarbon biodegradation and desorption rates in aged soils. *Biodegradation*, 15, 261–274. <https://doi.org/10.1023/B:BIOD.0000042996.03551.f4>
- Hughes, C. B., Brown, D. M., Camenzuli, L., Redman, A. D., Arey, J. S., Vione, D., Wang, N., & Vaipoulou, E. (2020). Can a chemical be both readily biodegradable AND very persistent (vP)? Weight-of-evidence determination demonstrates that phenanthrene is not persistent in the environment. *Environmental Sciences Europe*, 32, 1–19.
- Hur, J., Danes, L., Krout, D., Hsieh, J.-H., & Auerbach, S. (2017). Tox21 enricher: Web-based chemical and functional enrichment analysis tool for Tox21 toxicity screening platform. *The FASEB Journal*, 31, 613.1.
- Jaworska, J., Dimitrov, S., Nikolova, N., & Mekenyan, O. (2002). Probabilistic assessment of biodegradability based on metabolic pathways: CATABOL system. *SAR and QSAR in Environmental Research*, 13, 307–323.
- Kuhn, M., Weston, S., Keefer, C., & Coulter, N. (2012). Cubist models for regression. *R package Vignette R package version 0.0 18*.
- Kuo, D. T. F., & Di Toro, D. M. (2013). Biotransformation model of neutral and weakly polar organic compounds in fish incorporating internal partitioning. *Environmental Toxicology and Chemistry*, 32, 1873–1881. <https://doi.org/10.1002/etc.2259>
- Leonards, P. (2017). Concept for a generic information strategy for PBT and vPvB assessment of UVCB substances. Vrije Universiteit Amsterdam.
- Lofthus, S., Almås, I. K., Evans, P., Pelz, O., & Brakstad, O. G. (2018). Biodegradation in seawater of PAH and alkylphenols from produced water of a North Sea platform. *Chemosphere*, 206, 465–473.
- Lombardo, A., Pizzo, F., Benfenati, E., Manganaro, A., Ferrari, T., & Gini, G. (2014). A new in silico classification model for ready biodegradability, based on molecular fragments. *Chemosphere*, 108, 10–16.
- Mauri, A., Consonni, V., Pavan, M., & Todeschini, R. (2006). Dragon software: An easy approach to molecular descriptor calculations. *MATCH Communications in Mathematical and in Computer Chemistry*, 56, 237–248.
- McFarlin, K. M., Perkins, M. J., Field, J. A., & Leigh, M. B. (2018). Biodegradation of crude oil and Corexit 9500 in Arctic seawater. *Frontiers in Microbiology*, 9, 1788.
- McFarlin, K. M., Prince, R. C., Perkins, R., & Leigh, M. B. (2014). Biodegradation of dispersed oil in Arctic seawater at -1°C . *PLoS One*, 9, e84297. <https://doi.org/10.1371/journal.pone.0084297>

- Mekenyan, O., & Veith, G. (1994). The electronic factor in QSAR: MO-parameters, competing interactions, reactivity and toxicity. *SAR and QSAR in Environmental Research*, 2, 129–143.
- Meylan, W., Boethling, R., Aronson, D., Howard, P., & Tunkel, J. (2007). Chemical structure-based predictive model for methanogenic anaerobic biodegradation potential. *Environmental Toxicology and Chemistry: An International Journal*, 26, 1785–1792.
- Mo, Z. (2017). Overview of China new chemical substance notification (regulation framework, key point and implementation progress). *Journal of the Japan Society of Colour Material*, 90, 182–190.
- Mormile, M. R., Liu, S., & Sufflita, J. M. (1994). Anaerobic biodegradation of gasoline oxygenates: Extrapolation of information to multiple sites and redox conditions. *Environmental Science & Technology*, 28, 1727–1732.
- North, E. W., Adams, E. E., Thessen, A. E., Schlag, Z., He, R., Socolofsky, S. A., Masutani, S. M., & Peckham, S. D. (2015). The influence of droplet size and biodegradation on the transport of subsurface oil droplets during the Deepwater Horizon spill: A model sensitivity study. *Environmental Research Letters*, 10, 024016.
- Organisation for Economic Co-operation and Development. (2002a). Test No. 307: *Aerobic and anaerobic transformation in soil*.
- Organisation for Economic Co-operation and Development. (2002b). Test No. 308: *Aerobic and anaerobic transformation in aquatic sediment systems*.
- Organisation for Economic Co-operation and Development. (2004). Test No. 309: *Aerobic mineralisation in surface water—Simulation biodegradation test*.
- Organisation for Economic Co-operation and Development. (2019). Principles for the validation, for regulatory purposes, of (quantitative) structure–activity relationship models.
- Ossai, I. C., Ahmed, A., Hassan, A., & Hamid, F. S. (2020). Remediation of soil and water contaminated with petroleum hydrocarbon: A review. *Environmental Technology & Innovation*, 17, 100526.
- Pelletier, E., Delille, D., & Delille, B. (2004). Crude oil bioremediation in sub-Antarctic intertidal sediments: Chemistry and toxicity of oiled residues. *Marine Environmental Research*, 57, 311–327.
- Pizzo, F., Lombardo, A., Brandt, M., Manganaro, A., & Benfenati, E. (2016). A new integrated in silico strategy for the assessment and prioritization of persistence of chemicals under REACH. *Environment International*, 88, 250–260. <https://doi.org/10.1016/j.envint.2015.12.019>
- Prince, R. C., Butler, J. D., & Redman, A. D. (2017). The rate of crude oil biodegradation in the sea. *Environmental Science & Technology*, 51, 1278–1284.
- Prince, R. C., Haitmanek, C., & Lee, C. C. (2008). The primary aerobic biodegradation of biodiesel B20. *Chemosphere*, 71, 1446–1451.
- Prince, R. C., McFarlin, K. M., Butler, J. D., Febbo, E. J., Wang, F. C., & Nedwed, T. J. (2013). The primary biodegradation of dispersed crude oil in the sea. *Chemosphere*, 90, 521–526.
- Prince, R. C., Nash, G. W., & Hill, S. J. (2016). The biodegradation of crude oil in the deep ocean. *Marine Pollution Bulletin*, 111, 354–357.
- Prince, R. C., Parkerton, T. F., & Lee, C. (2007). The primary aerobic biodegradation of gasoline hydrocarbons. *Environmental Science & Technology*, 41, 3316–3321.
- Prosser, C. M., Redman, A. D., Prince, R. C., Paumen, M. L., Letinski, D. J., & Butler, J. D. (2016). Evaluating persistence of petroleum hydrocarbons in aerobic aqueous media. *Chemosphere*, 100, 542–549.
- Quinlan, J. R. (1993). Combining instance-based and model-based learning. In *Proceedings of the Tenth International Conference on Machine Learning* (pp. 236–243).
- Ren, X., Zeng, G., Tang, L., Wang, J., Wan, J., Liu, Y., Yu, J., Yi, H., Ye, S., & Deng, R. (2018). Sorption, transport and biodegradation—An insight into bioavailability of persistent organic pollutants in soil. *Science of the Total Environment*, 610–611, 1154–1163. <https://doi.org/10.1016/j.scitotenv.2017.08.089>
- Ribicic, D., McFarlin, K. M., Netzer, R., Brakstad, O. G., Winkler, A., Throne-Holst, M., & Størseth, T. R. (2018). Oil type and temperature dependent biodegradation dynamics—Combining chemical and microbial community data through multivariate analysis. *BMC Microbiology*, 18, 1–15.
- Ribicic, D., Netzer, R., Hazen, T. C., Techtmann, S. M., Drabløs, F., & Brakstad, O. G. (2018). Microbial community and metagenome dynamics during biodegradation of dispersed oil reveals potential key-players in cold Norwegian seawater. *Marine Pollution Bulletin*, 129, 370–378.
- Santas, R., Korda, A., Tenente, A., Buchholz, K., & Santas, P. (1999). Mesocosm assays of oil spill bioremediation with oleophilic fertilizers: Inipol, F1 or both? *Marine Pollution Bulletin*, 38, 44–48.
- Shrestha, P., Meisterjahn, B., Hughes, C. B., Mayer, P., Birch, H., & Hennecke, D. (2020). Biodegradation testing of volatile hydrophobic chemicals in water-sediment systems—Experimental developments and challenges. *Chemosphere*, 238, 124516.
- Siraki, A. G., Chevaldina, T., & O'Brien, P. J. (2005). Application of quantitative structure–toxicity relationships for acute NSAID cytotoxicity in rat hepatocytes. *Chemico-Biological Interactions*, 151, 177–191.
- Socolofsky, S. A., Gros, J., North, E., Boufadel, M. C., Parkerton, T. F., & Adams, E. E. (2019). The treatment of biodegradation in models of subsurface oil spills: A review and sensitivity study. *Marine Pollution Bulletin*, 143, 204–219.
- Spaulding, M. L. (2017). State of the art review and future directions in oil spill modeling. *Marine Pollution Bulletin*, 115, 7–19.
- U.S. Environmental Protection Agency. (2012). Work plan chemicals: Methods document. Office of Pollution Prevention and Toxics.
- U.S. Environmental Protection Agency. (2013). Interpretive assistance document for assessment of discrete organic chemicals. Office of Chemical Safety and Pollution Prevention.
- Veith, G. D., Mekenyan, O. G., Ankley, G. T., & Call, D. J. (1995). A QSAR analysis of substituent effects on the photoinduced acute toxicity of PAHs. *Chemosphere*, 30, 2129–2142.
- Voutchkova, A. M., Kostal, J., Steinfeld, J. B., Emerson, J. W., Brooks, B. W., Anastas, P., & Zimmerman, J. B. (2011). Towards rational molecular design: Derivation of property guidelines for reduced acute aquatic toxicity. *Green Chemistry*, 13, 2373–2379.
- Wang, Y., Lai, A., Latino, D., Fenner, K., & Helbling, D. E. (2018). Evaluating the environmental parameters that determine aerobic biodegradation half-lives of pesticides in soil with a multivariable approach. *Chemosphere*, 209, 430–438. <https://doi.org/10.1016/j.chemosphere.2018.06.077>
- Whale, G., Parsons, J., van Ginkel, K., Davenport, R., Vaiopoulou, E., Fenner, K., & Schaeffer, A. (2021). Improving our understanding of the environmental persistence of chemicals. *Integrated Environmental Assessment and Management*, 17, 1123–1135.
- Williams, A. J., Grulke, C. M., Edwards, J., McEachran, A. D., Mansouri, K., Baker, N. C., Patlewicz, G., Shah, I., Wambaugh, J. F., Judson, R. S., & Richard, A. M. (2017). The CompTox chemistry dashboard: A community data resource for environmental chemistry. *Journal of Cheminformatics*, 9, 61.
- Xu, R., & Obbard, J. P. (2003). Effect of nutrient amendments on indigenous hydrocarbon biodegradation in oil-contaminated beach sediments. *Journal of Environmental Quality*, 32, 1234–1243.
- Xu, R., & Obbard, J. P. (2004). Biodegradation of polycyclic aromatic hydrocarbons in oil-contaminated beach sediments treated with nutrient amendments. *Journal of Environmental Quality*, 33, 861–867.
- Yang, C., Tarkhov, A., Maruszczyk, J., Bienfait, B., Gasteiger, J., Kleinoeder, T., Magdziarz, T., Sacher, O., Schwab, C. H., Schwoebel, J., Terfloth, L., Arvidson, K., Richard, A., Worth, A., & Rathman, J. (2015). New publicly available chemical query language, CSRML, to support chemotype representations for application to data mining and modeling. *Journal of Chemical Information and Modeling*, 55, 510–528.
- Yap, C. W. (2011). PaDEL-descriptor: An open source software to calculate molecular descriptors and fingerprints. *Journal of Computational Chemistry*, 32, 1466–1474.