Environmental Toxicology and Chemistry—Volume 00, Number 00—pp. 1–12, 2024
Received: 26 September 2023 | Revised: 31 October 2023 | Accepted: 24 February 2024

Hazard/Risk Assessment

Predicting Hydrocarbon Primary Biodegradation in Soil and Sediment Systems Using System Parameterization and Machine Learning

Craig W. Davis,[a,*] David M. Brown,[b] Chesney Swansborough,[b] Christopher B. Hughes,[b] Louise Camenzuli,[c] Leslie J. Saunders,[d] and Delina Y. Lyon[d,*]

[a]ExxonMobil Biomedical Sciences, Annandale, New Jersey, USA
[b]Ricardo Energy & Environment, Harwell, UK
[c]ExxonMobil Petroleum & Chemical, Machelen, Belgium
[d]Concawe, Brussels, Belgium

Abstract: Technical complexity associated with biodegradation testing, particularly for substances of unknown or variable composition, complex reaction products, or biological materials (UVCB), necessitates the advancement of non-testing methods such as quantitative structure–property relationships (QSPRs). Models for describing the biodegradation of petroleum hydrocarbons (HCs) have been previously developed. A critical limitation of available models is their inability to capture the variability in biodegradation rates associated with variable test systems and environmental conditions. Recently, the Hydrocarbon Biodegradation System Integrated Model (HC-BioSIM) was developed to characterize the biodegradation of HCs in aquatic systems with the inclusion of key test system variables. The present study further expands the HC-BioSIM methodology to soil and sediment systems using a database of 2195 half-life (i.e., degradation time [DT]50) entries for HCs in soil and sediment. Relevance and reliability criteria were defined based on similarity to standard testing guidelines for biodegradation testing and applied to all entries in the database. The HC-BioSIM soil and sediment models significantly outperformed the existing biodegradation HC half-life (BioHCWin) and virtual evaluation of chemical properties and toxicities (VEGA) quantitative Mario Negri Institute for Pharmacological Research (IRFMN) models in soil and sediment. Average errors in predicted DT50s were reduced by up to 6.3- and 8.7-fold for soil and sediment, respectively. No significant bias as a function of HC class, carbon number, or test system parameters was observed. Model diagnostics demonstrated low variability in performance and high consistency of parameter usage/importance and rule structure, supporting the generalizability and stability of the models for application to external data sets. The HC-BioSIM provides improved accuracy of Persistence categorization, with correct classification rates of 83.9%, and 90.6% for soil and sediment, respectively, demonstrating a significant improvement over the existing BioHCWin (70.7% and 58.6%) and VEGA (59.5% and 18.5%) models. *Environ Toxicol Chem* 2024;00:1–12. © 2024 Concawe. *Environmental Toxicology and Chemistry* published by Wiley Periodicals LLC on behalf of SETAC.

Keywords: Biodegradation; Hazard/risk assessment; Quantitative structure–activity relationships

INTRODUCTION

The biodegradability of a chemical speaks directly to its ability to accumulate in the environment, or persist, and potentially cause unpredictable harm. This is an important aspect of the Stockholm Convention (United Nations Environment Programme, 2019) and multiple chemicals management regulations worldwide, such as the European Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH) regulation (European Union, 2006), the Plant Protection Products regulation (European Union, 2009), the Toxic Substances Control Act (US Environmental Protection Agency (USEPA), 1976), and the Canadian Environmental Protection Act (Environment and Climate Change Canada, 1999), with the objective to protect human health and the environment. Under these regulations, biodegradation assessment generally starts with a screening

---

This article includes online-only Supporting Information.

* Address correspondence to environment@concawe.eu and craig.w.davis@exxonmobil.com

DOI: 10.1002/etc.5857

assessment, such as the Organisation for Economic Co-operation and Development (OECD) series 301 of Ready Biodegradability tests (OECD, 1992). Chemicals that fail the ready biodegradability screening tests generally undergo higher tier testing in multiple environmental compartments including soil and sediment. The outcomes of the biodegradation tests are compared against Persistence criteria (Table 1) to determine whether further scrutiny of the chemical is warranted.

The objective of Persistence assessment is not to measure the intrinsic biodegradability of a chemical, that is, the ability of the chemical to be broken down by microbial processes, but to identify chemicals that should be prioritized for further assessment. Chemicals are identified as Persistent when they meet the regulatory half-life criteria, given the external factors that would amend the intrinsic biodegradability (Redman et al., 2021; Schäffer et al., 2022). Under REACH, according to the R.11 guidance (European Chemicals Agency [ECHA], 2023), these half-life values are calculated from standardized biodegradation simulation tests, specifically OECD test guidelines 307, 308, and 309, or from a weight-of-evidence assessment using available data that often include nonstandard data (ECHA, 2023). Simulation testing is often costly and not always feasible due to limitations in the applicability of the test guidelines for substances that are difficult to test (Redman et al., 2021; Shrestha et al., 2020). The REACH R.11 guidance (ECHA, 2023) defines a degradation half-life (DT50) as the time required for a test substance to reduce to 50% of its initial concentration in a test system in which the degradation can be described by (pseudo-) first-order reaction kinetics.

Half-lives derived from biodegradation tests can be highly variable, which is often attributed to the flexibility of the test set-up, allowing high variability in key external factors (Seller et al., 2021). These extrinsic factors that alter apparent biodegradability are difficult to standardize and are thus poorly accounted for in experimental test systems (Schäffer et al., 2022). Thus integration of these extrinsic factors/environmental properties to account for the differences in biodegradation seen in the different environmental media and across different tests is not yet well understood.

Extrapolation of biodegradation half-lives from one compartment to another is common practice in fate and exposure modeling. These intermedia extrapolation factors (IMEFs) provide a quick way to estimate biodegradation half-lives in compartments for which no data are available. In 1995, Boethling et al. suggested factors of 1:1:4 for extrapolation between water, soil, and sediment systems based on relative rates of biodegradation in the respective environmental compartments (Boethling et al., 1995). This was revised in 2006 to 1:2:9, as incorporated in the USEPA EPI Suite model (Aronson et al., 2006). However, these IMEFs are rough estimations based on highly variable test data or on expert judgment. Hence there remains a need for methods to obtain biodegradation data for the soil and sediment compartments beyond simulation testing, particularly for difficult to test substances.

Another option for Persistence screening in soil and sediment is the use of quantitative structure–property relationships (QSPRs) for biodegradation based on empirical data. Such QSPR models use structural, chemical, or environmental properties to develop a quantitative relationship for predicting a given property of interest (in this case, biodegradation rates in the environment). They are similar to quantitative activity–structure relationships (QSARs), in which the specific "property" used is an activity (e.g., $\log(K_{OW})$ to predict membrane–water partitioning; Endo & Goss, 2014). Numerous biodegradation QSPRs exist for aqueous systems (Pavan & Worth, 2006; Singh et al., 2021), but there are far fewer available for soil and sediment systems. The outputs of the QSPRs can range from indications of "ready biodegradability" to calculations of biodegradation half-lives. Many biodegradation QSPRs generate semiquantitative biodegradation half-lives for water, soil, and sediment, and are generally based solely on molecular structure (Lombardo et al., 2022). However, recent work by Davis et al. (2022) has shown that molecular structure information alone is not sufficient to reliably predict the primary biodegradation of hydrocarbons (HCs), and that test system parameters play a key role in biodegradation estimation for aqueous systems. This assessment informed the development and the application of the Hydrocarbon Biodegradation System Integrated Model (HC-BioSIM) to predict biodegradation in freshwater and seawater test systems (Davis et al., 2022). In more complex soil and sediment systems, it is expected that test system parameters will similarly play a critical role in the prediction of the biodegradation. For example, in one soil biodegradation QSPR model for phthalate esters, soil pH and organic matter content were recognized as soil properties

TABLE 1: Persistent and very persistent criteria under REACH, TSCA and CEPA

	Half-life criteria		
Persistence	REACH	TSCA	CEPA
Persistent (P)	>60 days in marine water >40 days in fresh or estuarine water >180 in marine sediment >120 days in fresh or estuarine sediment >120 days in soil	≥60 days in water, sediment, and soil ≥2 days in air	≥2 days in air ≥182 days in water ≥365 days in sediment ≥182 days in soil
Very Persistent (vP)	>60 days in marine, fresh, or estuarine water >180 days in marine, fresh, or estuarine sediment >180 days in soil	>180 days in water, soil, and sediment	—

REACH = Registration, Evaluation, Authorisation and Restriction of Chemicals regulation; TSCA = Toxic Substances Control Act; CEPA = Canadian Environmental Protection Act.

impacting chemical partitioning and thereby impacting predictions of phthalate biodegradation rates (Bai et al., 2021).

As stated previously, Persistence assessment under REACH requires biodegradation half-lives for all relevant environmental compartments (water, soil, and sediment), which could be generated by QSPR (REACH Annex XI). The aims of our study were (1) to expand a recently constructed database of aerobic HC biodegradation half-lives in water from Brown et al. (2020) to include aerobic HC biodegradation in soil and sediment; and (2) to generate an HC-specific soil and sediment biodegradation QSPR including system parameters, building off of the HC-BioSIM water model (Brown et al., 2020; Davis et al., 2022).

MATERIALS AND METHODS

Database collation for HC biodegradation in soil and sediment systems

The HC biodegradation half-life database was compiled using aerobic biodegradation half-life data from peer-reviewed literature, obtained predominantly through a systematic search of scientific literature repositories (i.e., PubMed, Science Direct, Google Scholar, Deepdyve) using relevant keywords (Supporting Information Database SI, Table S3). Gray literature was also identified using relevant keyword searches of Google and the reference lists of known HC biodegradation databases, such as the USEPA's biodegradation HC half-life (BioHCWin) and Aerobic Biodegradation of Organic Chemicals (Aronson et al., 1999; European Centre for Ecotoxicology and Toxicology of Chemicals, 2009; Howard et al., 2005). The full database is available as an Excel sheet in the Supporting Information titled "Database SI."

A title and abstract screening process was undertaken, followed by a second round of review of the full texts. The following quality screening criteria were applied for selection of studies to be included in the database: (1) aerobic biodegradation data derived from studies using naturally occurring inocula from uncontaminated environments that have no direct or obvious pre-exposure to HCs; (2) studies conducted with freshly spiked HC, hence excluding any data influenced by aging with the substance; (3) studies involving measurement of primary biodegradation of individual HC constituents; and (4) studies that minimized or accounted for abiotic losses.

A range of key study parameters and metadata were defined (the full list is available in the Supporting Information, Table S4). These covered aspects such as the identity and form of the test substance, inoculum characteristics, experimental conditions, and results, and they were used to inform the assessment of the study quality. Biodegradation kinetics are a key metric captured as biodegradation half-lives (DT50) and are described further in the *Biodegradation kinetics* section.

Data quality criteria were developed to assess the studies in the database and support their use for regulatory purposes and other related uses, including QSPR development. Relevance and reliability criteria were defined based on similarity to standard OECD test guidelines, and other recent work (Wassenaar & Verbruggen, 2021). Relevance criteria focused on the characteristics and treatment of the inoculum and experimental test set-up. Reliability criteria centered around the use of controls, dosing strategies, and replicates. Further reliability criteria were developed for assessment of the measurement of biodegradation. The full list of criteria for relevance and reliability and their justification is available in the Supporting Information, Tables S5 and S6, respectively, with an overview of the selected parameters shown in Table 2. The full range of reliability and relevance parameters was evaluated for each individual study in the database. Low, medium, and high rankings were assigned scores of 1, 2, or 3, respectively, to allow data to be scored against best practices. Mean average

TABLE 2: Overview of the relevance and reliability criteria used for database development

	Relevance parameter		Reliability parameter
Test system	Water:sediment ratio in test system ($R_{W:S}$; v/v)	Test system	Dosing strategy, i.e., approaches taken to ensure homogeneous distribution and bioavailability of test substance
	Soil moisture content during study		
	Sediment agitation		
	Aeration of test system		Measures to assess or minimize abiotic losses
	Light regime during study		
	Test substance concentration[a]; (dose mg/kg; ppm)		
	Application of test substance as crude oil or a complex mixture		No. of study replicates
	Test temperature (T; °C)		
	Experimental setup		
Inoculum	Sediment characteristics (pH, organic carbon content (f_{oc}; %)	Measurement	No. of sampling time points during the study
	Soil characteristics (soil classification, pH, f_{oc})		Test substance recovery, i.e., chemical extraction efficiency and mass balance
	Sampling and storage of soil and sediment		Analytical method reliability
	Supplementation with nutrients or mineral media		Extent of biodegradation observed during study and used for HL/DT50 calculation
	Type of inoculum used		HL/DT50 calculation

[a] For the purposes of our study, loading refers to the total hydrocarbon initial concentration, and dose refers to the initial concentration of the individual test substance(s) (for which DT50 values are reported).

Additional information related to the relevance and reliability parameters, criteria, and justifications is available in the Supporting Information, Tables S5 and S6.

HL/DT50 = half-life excluding lag phase/half-life + lag phase.

scores were calculated for each category by averaging the two category scores for each metric. All individual parameters were assigned equal weightings.

Biodegradation kinetics

The biodegradation data found in studies were presented in a number of different forms. For the purpose of our study, the convention established concerning biodegradation endpoints in Birch et al. (2018), and adopted in Brown et al. (2020) and Davis et al. (2022), was followed (Birch et al., 2018; Brown et al., 2020; Davis et al., 2022). Briefly, when degradation data permitted a lag phase (period of no discernible degradation) to be distinguished from the degradation phase, this lag phase was captured and reported separately. In the database, separate degradation endpoints are presented including and excluding any lag phase, referred to as either half-life (excluding lag phase) or DT50 (half-life + lag phase). We have defined DT50 as a term for use in our study, but the reader should verify whether our definition aligns with the definition for their particular regulatory context.

Data already in the form of a half-life were directly input into the database with no further treatment. When biodegradation rates (k; days^{-1}) were available, half-lives were calculated from the rate information according to Equation (1). Degradation graphs or time series data were assessed to obtain concentration and time data, which were then used to calculate half-life and DT50 (both in units of days). When it was not possible to obtain data from graphs (i.e., due to low resolution), half-life was determined directly by visual inspection of the graph.

$$\text{Half-life} = \frac{\ln(2)}{k} \tag{1}$$

Two kinetic approaches were applied to calculate DT50 values from biodegradation time series data for inclusion in the database. The first describes a "one-phase decay model from a plateau," incorporating pseudo-first-order kinetics with an initial lag phase (X_0; days). For studies in which a lag period was clear (degradation not greater than 10%) and overall degradation during the study period was greater than 50%, this approach was applied. The kinetics for this model were determined using the "plateau + exponential decay model" on Prism 9 (GraphPad, 2020). This equation calculates the two phases as distinct, separate kinetic components and is described as

$$\text{DT50} = X_0 + \frac{\ln(2)}{k} \tag{2}$$

Alternatively, some DT50 values incorporated 0-order degradation kinetics with an initial lag phase. Using $[A]$ as the current concentration and $[A]$o as the initial concentration, k is the reaction constant, t is time in the equation $[A] = [\underline{A}]_0 - kt$, the DT50 is calculated as follows:

$$\text{DT50} = X_0 + \frac{[A]_0}{2k} \tag{3}$$

Model development

The compiled databases (soil and sediment) as just described were used to develop and validate the respective HC-BioSIM QSPR models. Previously, a decision-tree machine learning model, HC-BioSIM, had been developed for predicting primary biodegradation disappearance times (DT50) of HCs in marine and freshwater systems (Davis et al., 2022). Briefly, the model utilizes a decision-tree algorithm to develop a series of rules from a set of chemical structure and system parameters (Quinlan, 1993). These rules parse and subset the DT50 data, maximizing similarity within the subsets and minimizing the entropy of the entire system. For each subset, a multiple linear regression is applied to generate quantitative DT50 predictions, drawing from the pool of remaining structural and system parameters. Redundant rules are trimmed or combined in a postdevelopment "pruning" process to minimize model complexity and limit overparameterization. A generalized schematic of the model architecture is presented alongside the model output (Supporting Information Sediment Model and Soil Model, Table S1). This approach results in a model architecture and output that (1) is transparent, (2) is easily communicated, and (3) provides a mechanistic basis for interpretation and application, in line with the OECD principles for the development and validation of QSAR/QSPR models (OECD, 2014).

ToxPrint structural fingerprint parameters (Yang et al., 2015) were selected to describe key structural features of the constituents in the soil and sediment database. ToxPrint fingerprints were obtained from the USEPA's CompTox Dashboard (Williams et al., 2017) and were further manually curated to eliminate redundant structural descriptors. This is consistent with the previously developed HC-BioSIM water model (Davis et al., 2022). For the soil model, test temperature (T), test substance concentration (Dose), and fraction organic carbon (f_{oc}) were identified as relevant system parameters because they have been previously identified as influencing the biodegradation of HCs in soil (Davenport et al., 2022). For the sediment model, one additional parameter, the ratio of water to sediment ($R_{W:S}$) was included as a relevant system parameter in lieu of f_{oc}, because the $R_{W:S}$ can be considered a surrogate for the relative amount of organic carbon in the test system (in the absence of direct reported f_{oc} data).

Chemical and system parameters were then combined to create a single pool of parameters for the development and calibration of the soil and sediment models. To systematically evaluate the influence of structure and system parameters, a step-wise approach to model construction was implemented. As a base case, a "structure-only" model was developed that provides a consistent comparison to the existing BioHCWin and VEGA models (Howard et al., 2005; Lombardo et al., 2022). System parameters were then sequentially added, with performance (root mean square error [RMSE] and R^2, discussed in the following section, *Model performance and data analysis*) and complexity (number of rules and parameters used in the models) evaluated for each subsequent iteration of the model. Additional consideration was given to parameters that may have a specific impact on regulatory decision-making or relevance of the predicted value for comparative assessment (e.g., comparing a model result with OECD simulation study conditions). Finally, a model was selected that balanced complexity, performance, and end-user application needs. All information is available in Table S2 of the Supporting Information for each model (soil and sediment). Model calibration, cross-validation, and statistical analyses were performed in R (Ver. 3.6.1) unless otherwise stated. The R-code files for the soil and sediment models (including parameter curation and visualizations) are available as R script in the Supporting Information.

Model performance and data analysis

Soil and sediment models were trained and validated using the curated experimental database discussed previously. Random seeds were established to ensure unbiased and reproducible selection of training (80%) and validation (20%) subsets as well as for the fold selection in the *k*-fold cross-validation analysis. Model performance was compared with predictions made with the existing BioHCWin model (using 1:1:4 water:soil:sediment IMEF; Boethling et al., 1995; Howard et al., 2005) as well as the VEGA persistence soil and sediment models (Benfenati et al., 2013; Lombardo et al., 2022). The BioHCWin model was selected because it is used routinely in chemical registration and hazard assessment and is freely available within the USEPA's EPI Suite software package (USEPA, 2023). The VEGA models were included because they are the only quantitative models for soil and sediment compartments that are (1) applicable to a wide set of chemicals, (2) free to use, and (3) easily accessible (i.e., via a web platform; Benfenati et al., 2013; Lombardo et al., 2022). Prediction accuracy was evaluated using the RMSE as well as the Pearson correlation coefficient (R^2). The RMSE values for the model predictions are computed as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(\log(DT50_{\exp,i}) - \log(DT50_{pred,i}))^2}{N}} \tag{4}$$

where N is the number of observations, $DT50_{\exp,i}$ is the experimental DT50 value of the *i*th observation, and $DT50_{pred,i}$ is the predicted DT50 value. Logarithmic DT50 values are used in the calculation of the RMSE for two reasons: first, to assign equal weighting across a large numerical range of observed DT50 values; and second, to provide a metric that is easily interpreted and communicated. For example, RMSE values of 0.3 and 0.5 log units correspond to average error of predictions of 2*x* and 3*x*, respectively.

Finally, a *k*-fold cross-validation ($k = 5$) was performed on the soil and sediment models. Mean and standard deviations of RMSE and R^2 values, as well as concordance between the rules structure and parameter usage were assessed to provide an overall assessment of the stability and generalizability of the calibrated models. The results are available in the Supporting Information, with a complete description of the cross-validation process (and associated diagnostics) provided in the Supporting Information of Davis et al. (2022) in line with the OECD guidance on QSAR/QSPR development (OECD, 2014).

RESULTS AND DISCUSSION

Soil and sediment database

Following the literature search and screening detailed above, a total of 100 independent studies (58 soil studies and 42 sediment studies) containing suitable biodegradation test data were identified. Of the 221 studies available in the BioHCWin database, only 34 studies were found to be suitable for inclusion in the soil and sediment database. Reasons for exclusion included inappropriate environmental inoculum (water), pre-exposure of the inoculum, and incorrect endpoint measurement (mineralization).

The final curated HC biodegradation database consisted of 2195 entries (1243 in sediment and 952 in soil). All studies meeting the selection criteria are listed in Table S2 of the Database SI Supporting Information. Studies that failed to meet the selection criteria are listed in the Database SI Supporting Information, Table S7 (omitted studies), along with the reason for their omission. Almost 50% of the studies omitted from the final database were omitted due to pre-exposure or adaptation to the test substance, which is not permitted according to REACH R.11 guidance (ECHA, 2023). Other reasons included the measurement of ultimate biodegradation and the HC constituents not being described. Of the 2195 half-lives compiled, 16 sediment half-lives and 4 soil half-lives were excluded from the model training and validation data set because the computed half-lives reported were greater than the duration of the study and were considered to be unreliable for the purpose of calibrating a quantitative model. This is consistent with the previous development of the aquatic DT50 database (Davis et al., 2022).

Figure 1 illustrates the distribution of chemical class and carbon number (CN) in soil (left) and sediment (right) studies across the database. Data covering more than 170 individual HC constituents were compiled in the database and cover all major HC chemical classes. Linear alkanes (n-paraffins), monoaromatic hydrocarbons (MAHs) and polyaromatic hydrocarbons (PAHs) represent the majority of the database. Carbon number ranged from 3 to 66, but most of the data were from constituents with $CN < 30$ in both sediment and soil studies. At least one half-life for *n*-paraffins covering every CN from 3 to 66 was included in the database. The PAH data were mostly available for HCs with CNs between 14 and 18, whereas the majority of the data for MAHs and di-aromatic hydrocarbons (DAHs) was for constituents with CNs $\leq$ 12. The only data for constituents with $CN > 23$ were for DAHs and *n*-paraffins.

Data quality evaluation

The data relevance and reliability criteria (Database Supporting Information, Tables S5 and S6) were applied to all entries in the database, and the distribution of overall scores is shown in Figure 2. There was no correlation between assigned

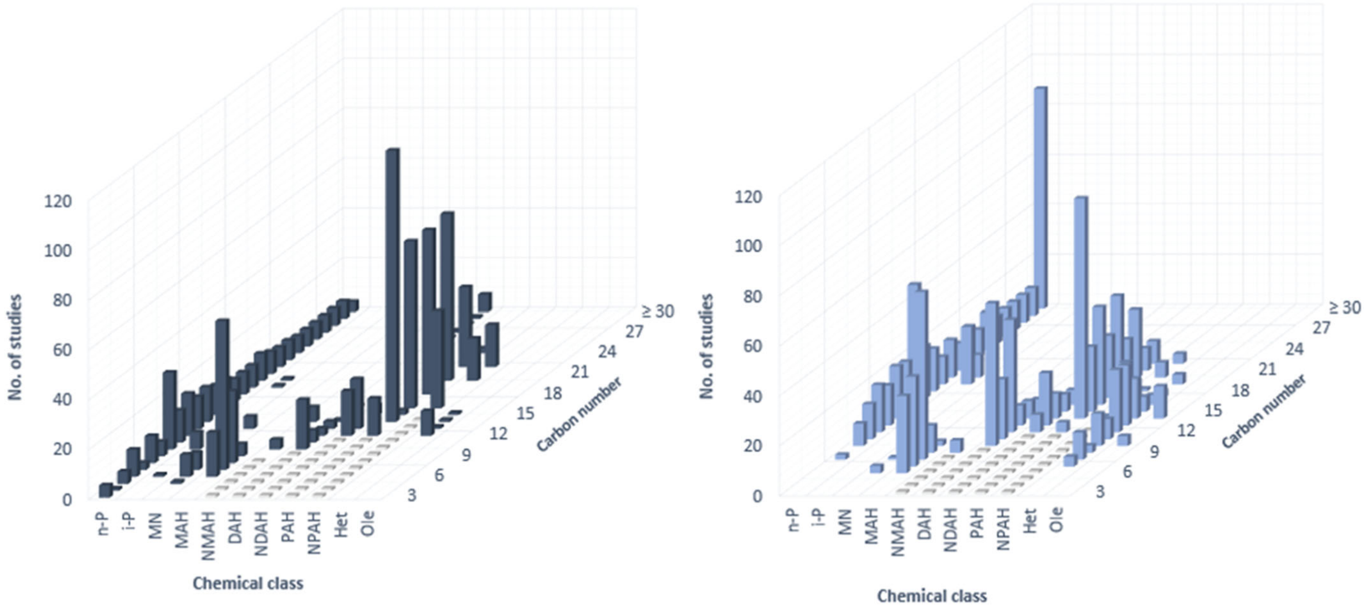

FIGURE 1: Distribution of chemical class and carbon number across the database in soil (left) and sediment (right), with normal paraffin (nP); isoparaffin (i-P); mono-naphthenic (MN); olefin (Ole); mono-aromatic (MAH); mononaphtho-monoaromatic (NMAH); diaromatic (DAH; including alkylated forms); naphthenic-diaromatic (NDAH); polyaromatic (PAH); naphthenic-polyaromatic (NPAH); and heterocyclic hydrocarbon (Het). Boxes have been grayed out where no possible structures exist for the corresponding carbon number.

relevance and reliability scores, as assessed by the low R^2 value (0.1025) of a plot of the relevance versus reliability scores.

Out of the 2195 half-lives assessed, 1051 (47.9%) were flagged as possible critical fails for incubation in the presence of light, 121 (5.5%) for inhibitory effects on the inoculum, 8 (0.4%) for possible anaerobic conditions, and 124 (5.6%) for potential bioavailability issues. When time and/or concentration values were not available and needed to be derived from a study graph image, 1117 (50.7%) data points were flagged. In total, <15% of data points were not assigned any quality flags.

The assessed parameters with the highest percentages of low relevance scores were organic carbon content (62.7%), water:sediment ratio (65.7%), pH (57.2%), and sediment agitation and aeration (56.8%). Seven hundred twenty-five (32.9%) data points achieved an average relevance score of ≥2.5, 1480 (67.1%) scored between 1.5 and 2.5, and no entry scored <1.5.

Generally, the majority of data points assessed scored high for reliability criteria, with the exceptions of test substance recovery, analytical method reliability, and half-life/DT50 calculation. Only 1.1% and 6.0% of all data points were assigned high scores for half-life/DT50 calculation and analytical method reliability, respectively. The majority were assigned medium scores for these categories. Only 6 averaged a reliability score of <1.5 (0.3%), whereas 854 (39%) scored between 1.5 and 2.5, and 1335 (61%) scored ≥2.5.

Model performance

The comparison of predicted and experimental DT50 values using the final selected soil and sediment models is presented in Figure 3. Model performance was benchmarked against the existing aqueous BioHCWin model with a 1:1:4

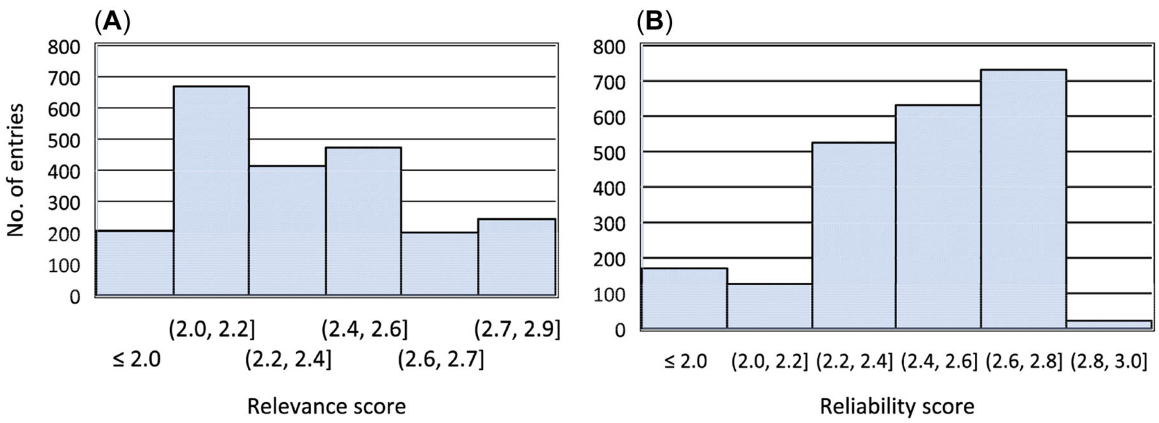

FIGURE 2: Distribution of overall (**A**) relevance and (**B**) reliability scores for all entries in the database.

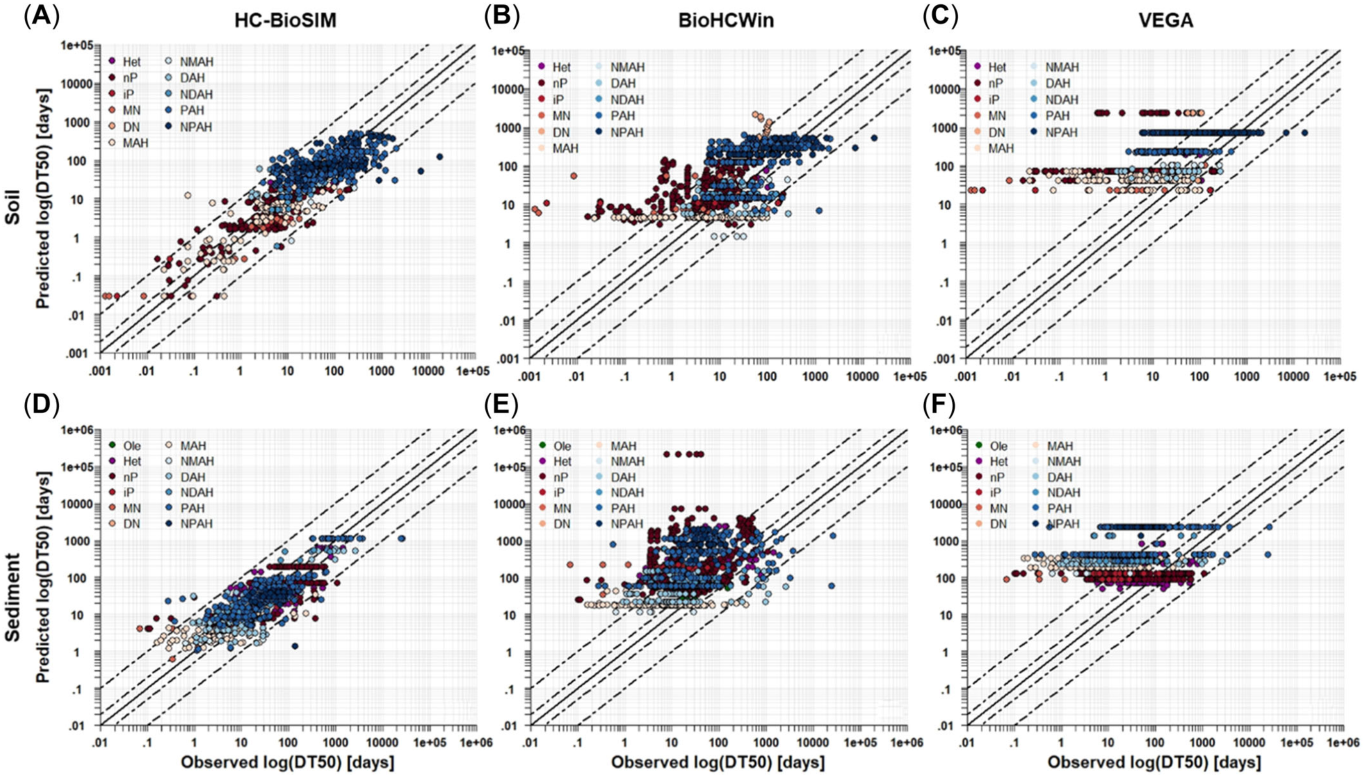

FIGURE 3: Comparison of the Hydrocarbon Biodegradation System Integrated Model (HC-BioSIM), BioHCWin, and VEGA QSPR models (predicted vs. experimental disappearance times [DT50; in days]) for soil (**A**–**C**) and sediment (**D**–**F**) data sets. Solid lines represent 1:1 agreement between predicted and experimental values. Dot-dashed lines represent 2x and 3x average errors (root mean square error = 0.3 and 0.5), respectively. Colors correspond to hydrocarbon classes: normal paraffins (nP), iso-paraffins (iP), mono-naphthenics (MN), di-naphthenics (DN), mono-aromatics (MAH), di-aromatics (DAH), poly-aromatics (PAH), naphthenic mono-aromatics (NMAH), naphthenic di-aromatics (NDAH), naphthenic poly-aromatics (NPAH), olefins (Ole), and S-, N-bearing heteroatoms (Het).

water:soil:sediment IMEF (Boethling et al., 1995; Howard et al., 2005) as well as directly against available quantitative models for soil and sediment half-lives from the VEGA model platform (Benfenati et al., 2013; Howard et al., 2005). Individual predictions for these models are available in the respective models in the Sediment Model and Soil Model Supporting Information, Table S3. Summary statistics (i.e., RMSE and R^2) are provided for the overall soil and sediment data sets as well as separately for the training and validation sets (Table 3). A complete analysis of HC-BioSIM residuals including boxplots for system parameters, HC chemical class, and CN is available in Table S5 of the Soil Model or Sediment Model Supporting Information (residuals).

Soil model

For soil, improvements over BioHCWin and VEGA were observed for the "base" structure-only HC-BioSIM model, with a 1.5–4*x* reduction in RMSE (see the Supporting Information, Stepwise Model Investigation). Even so, improvements in R^2 over the existing models were small (0.10–0.15), highlighting the need for inclusion of additional system parameters. Inclusion of the log (f_{oc}) as a system-parameter produced the most substantial performance increase, with a 2- to 6-fold reduction in RMSE and a significant increase in the ability of the HC-BioSIM model to capture the variability in the observed DT50 values ($R^2 = 0.74$). Less significant model improvements were observed with the single addition of either Dose or Temperature parameters, respectively. Improvements after addition of a second and/or third system parameter were less significant but did not increase the overall complexity of the model(s). These results are expected, because organic carbon in a soil system will alter both the bioavailability of an HC substance and microbial abundance in the test system. Thus, f_{oc} combined with the dose is likely to have a significant modifying effect on observed biodegradation half-lives in the system (Posada-Baquero et al., 2024; Shrestha et al., 2019). This is reflected in both the relative importance of the added system parameters and the usage in constructing the model rules/architecture. The final soil model included f_{oc}, dose, and temperature. Ultimately, temperature was included because it is a key factor in enabling regulatory Persistence assessment as well as important for providing a valid operational domain for the use of the model in other settings (e.g., oil spill modeling). The complete stepwise model evaluation, the resulting performance and rule(s) parameters, and a k-fold cross-validation of the final model(s) are provided in the Supporting Information, Stepwise Model Investigation.

The final HC-BioSIM soil model reduced the average DT50 prediction error (RMSE) by 0.4 to 0.8 log-units (i.e., by a factor of 2.5–6.3x) compared with the BioHCWin and VEGA models, respectively. Performance was similar between the training (RMSE = 0.46, *N* = 756) and the validation (RMSE = 0.54,

TABLE 3: Summary statistics (root mean square error [RMSE] and R^2) for the overall soil and sediment data sets as well as for the training, validation, and total data sets

			HC-BioSIM		BioHCWin		VEGA	
System	Data set	No.	RMSE	R^2	RMSE	R^2	RMSE	R^2
Soil	Training	756	0.438	0.78	0.847	0.37	1.240	0.33
	Validation	190	0.542	0.71	0.848	0.46	1.271	0.37
	All DT50	946	0.459	0.76	0.847	0.39	1.246	0.34
Sediment	Training	986	0.402	0.71	1.160	0.17	1.36	0.04
	Validation	247	0.414	0.58	1.090	0.09	1.26	0.02
	All DT50	1233	0.405	0.68	1.150	0.16	1.34	0.04

DT50 = half-life + lag phase; HC-BioSIM = Hydrocarbon Biodegradation System Integrated Model.

$N = 190$) data sets, as well as between the folds of the *k*-fold validation analysis (data available in the Supporting Information Soil Model, Table S4). This strongly suggests a stable solution, a model that is highly generalizable, and a model whose structure and performance are relatively insensitive to the selection of training data. Finally, no significant predictive bias was observed across the soil data set, either as a function of HC chemistry (CN, chemical class) or the parameters of the test system (f_{oc}, Dose, Temperature). This indicates that the selection of parameters, as well as the model architecture, are unbiased in characterizing the influence of the test system as well as the HC structure on the observed DT50. Box plots of the computed residual errors (Equation 4) for HC structure and test system parameters are presented in the Supporting Information Soil model, Table S4 Residuals).

An obvious limitation of the existing BioHCWin and VEGA model architecture can be observed in Figure 3B and C. Horizontal striations of predicted values for naphthenic-aromatic structures (naphthenic mono-aromatics, naphthenic di-aromatics, naphthenic poly-aromatics NPAH) as well as some higher aromatic compounds (PAHs) in the BioHCWin model and for all HC classes in the VEGA model suggest that neither a linear fragment-addition approach nor an approach based solely on chemical structural alerts are adequate to represent the DT50 behavior for these constituents. For example, reported DT50 values for benzo(*g*,*h*,*i*)perylene ranged from 40 to approximately 17,000 days, with a geometric mean value of 355 days. These studies have experimental conditions that varied considerably (T: 10–30 °C, f_{oc}: 0.5%–11%, dose: 0.7–184.5 mg/kg-soil wet wt). The single BioHCWin-predicted value of 517 days, although able to reproduce the average experimental result, failed to capture the substantial experimental variability across the various studies and test systems. The inclusion of the test system parameters in HC-BioSIM allows for a marked improvement in the ability to reliably and quantitatively capture this experimental variability.

Sediment model

For sediment, a more significant improvement in RMSE over the existing BioHCWin and VEGA models was observed for the "structure only" base case, with a 3*x* and 5*x* reduction in error, respectively (see the Supporting Information, Stepwise Model Investigation). However, as with soil, R^2 values for the base models were not significantly improved. For the sediment model, log(f_{oc}) was removed as a system parameter due to the high fraction of DT50 values that did not report f_{oc} values for test sediments (88%). Inclusion of the first system parameter produced the most dramatic improvement in model performance, with dose, $R_{W:S}$, and temperature providing the most substantial improvements, respectively. The additional sediment parameter ($R_{W:S}$) can be considered mechanistically a surrogate for the partitioning of the substance in the sediment system in the absence of f_{oc} data (a critical parameter in the soil model) and has been shown previously to be a crucial factor in sediment biodegradation simulation tests (Honti & Fenner, 2015). It is possible that the inclusion of direct f_{oc} data for sediment systems may provide additional information and should be a focus of future work. As with the soil model, RMSE improvements with the addition of a second and third system parameter were small and did not result in increased model complexity. It should be noted as well that the inclusion of 2+ and 3+ ring PAHs and long alkyl chain length structural parameters as minor modifying rule parameters was consistent between the soil and sediment models as well as with the previously developed water model (Davis et al., 2022). The final sediment model included $R_{W:S}$, dose, and temperature. As for soil, temperature was included to address explicitly the relevance and domain of the model outputs for regulatory and other applications. The complete stepwise model evaluation, the resulting performance and rule(s) parameters, and a *k*-fold cross-validation of the final model(s) are provided in the Supporting Information (Stepwise Model Investigation).

The final HC-BioSIM sediment model reduced the average DT50 prediction error (RMSE) by factors of 5.6 and 8.7*x* compared with the BioHCWin and VEGA models, respectively. In the case of BioHCWin, this enhanced improvement is likely due in part to the selection of the water:sediment IMEF. A standard IMEF value of 1:4 was selected (Boethling et al., 1995). It is clear from Figure 3E that systematic overprediction of the DT50 values occurs across the entire data set, suggesting that the 1:4 IMEF is overly conservative for HC constituents. This observation is important, because often even larger extrapolation factors (e.g., 1:9 [Aronson et al., 2006]) are suggested for Persistence assessment in sediments. This result suggests that employing existing IMEFs may greatly overestimate the DT50 in sediment systems for HCs. Performance for the training (RMSE = 0.39, $n = 986$) and validation (RMSE = 0.42, $n = 247$) sets, as well as the *k*-fold cross-validation (Supporting Information, Sediment model, Table S4) were consistent, again supporting the stability and generalizability of the HC-BioSIM framework.

It is interesting to note that the HC-BioSIM sediment model failed to discriminate a range of experimental DT50 values for several higher carbon range normal paraffin structures (nP) as well as several PAH compounds, including C1–C2 chrysenes/pyrenes, C4 phenanthrenes, and several phenanthrene, pyrene, and chrysene datapoints (Figure 3D). On further inspection, the nP data were from a single study, in which all system parameters were held constant except for the application of dispersant (Bacosa et al., 2018). That study did not report a log(f_{oc}) value for the test sediment. Predicted log(K_{ow}) values (as a surrogate for organic carbon–water partitioning) were computed from EpiSuite (Ver. 4.11) and plotted against the observed log(DT50) values for both the dispersed and nondispersed compounds. A strong linear relationship was observed. For the PAH compounds, the range of log(K_{ow}) values is small (~5–7), and no strong trend was observed (see the Supporting Information for raw data).

Because there is a large variability in the sediment f_{oc} in a test system and in substance hydrophobicity (K_{ow}), a considerable uncertainty in the regulatory application of model predictions may exist when the sediment f_{oc} is not reported. This is particularly likely for constituents that sorb strongly to organic carbon (e.g., those with high organic-carbon partition coefficients [K_{oc}]). In future analyses, it may be prudent to flag DT50 predictions when information on the test system is unknown. Even so, no strong predictive bias was observed across the sediment data set, either as a function of HC chemistry (i.e., CN, chemical class) or the remaining system parameters (i.e., temperature, dose). This lack of bias indicates that the selection of parameters as well as the model architecture is adequate to characterize the influence of the test system as well as the HC structure on the observed DT50. Box plots of the residual errors (Equation 4) for HC structure and test system parameters are available in the Supporting Information, Sediment model, Table S5 (Residuals).

Persistence categorization

In addition to the quantitative predictive capability, it is important to understand the ability of the model to accurately categorize the Persistence of the HCs under the varying test conditions. This can improve our understanding of how test conditions can potentially lead to conflicting or erroneous Persistence conclusions and can also highlight test conditions or systems that may not be representative of relevant environmental systems. To evaluate their respective performance, paired predicted and experimental DT50 values were compared against the REACH persistent (P) and very persistent (vP) criteria (Table 1) in soil or sediment. The percentage of correct predictions, and the false-positive and false-negative conclusions relative to a Persistence conclusion based on an experimental DT50 value were compared for the HC-BioSIM, BioHCWin, and VEGA models. The summary results of this analysis are presented in Table 4, and the full results are available in the Supporting Information for the soil and the sediment models, Table S6 (P categorization).

Significant improvement in correct prediction was observed broadly for HC-BioSIM for both soil and sediment systems. Accuracy of Persistence categorization improved by 10% to 23% for soil and 32% to 72% for sediment predictions relative to the BioHCWin and VEGA models. This largely corresponds to the correction of false-positive predictions (–10% and –30%, for BioHCWin and VEGA models, respectively). However, due to the increase in accuracy, HC-BioSIM appears to be less conservative than the other models, with an increase in false-negative predictions for the HC-BioSIM soil predictions (12.7%) relative to BioHCWin (2.2%) and VEGA (0.8%). However, these false negatives corresponded to DAH, PAH, and NPAH constituents tested as mixtures with total HC concentrations (loading) that can be considered relatively high; reported concentrations ranged from 6 to 172,000 mg/kg sediment with a geometric mean value of 198 mg/kg sediment. The combination of high test concentrations and affinity for binding to organic carbon may reduce the bioavailability of the individual constituents, and result in slower than anticipated biodegradation. If studies with total HC concentrations >100 mg/kg are removed, the recomputed false-negative rating for the HC-BioSIM soil model is 5.9%, much closer to those of the BioHCWin and VEGA models. For the sediment DT50 predictions, the percentage of false-negative ratings for each model was relatively similar (6.7% [HC-BioSIM], 4.7% [BioHCWin], 6.1%, [VEGA]). Maintaining a low rate of false-negative categorization while substantially reducing false-positive categorization is important, because it demonstrates the ability of HC-BioSIM to be used reliably as an initial screening and

TABLE 4: Comparison of persistence conclusions (%; using the REACH P and vP criteria in Table 1) based on DT50 values predicted by HC-BioSIM, BioHCWin, or VEGA models versus experimental DT50 data

	Predicted persistence compared with persistence conclusion based on empirical data	HC-BioSIM	BioHCWin	VEGA
Soil	Correct prediction	83.9	70.7	59.5
	Overprediction (false positive)	3.4	26.2	39.6
	Underprediction (false negative)	12.7	2.2	0.8
Sediment	Correct prediction	90.6	58.6	18.5
	Overprediction (false positive)	2.8	36.7	75.4
	Underprediction (false negative)	6.7	4.7	6.1

DT50 = half-life + lag phase; HC-BioSIM = Hydrocarbon Biodegradation System Integrated Model; P = persistent; REACH = Registration, Evaluation, Authorisation and Restriction of Chemicals regulation; vP = very persistent.

prioritization tool for HCs across a wide range of soil and sediment test conditions.

Model domain and application to OECD 307/308 systems

As discussed in the *Soil and sediment database* section, the soil and sediment data sets that we compiled include a broad range of experimental designs, test systems, and environmental test parameters. Nonstandard degradation experiments differ in their experimental setups from standard OECD tests, which limits their regulatory acceptance to a weight-of-evidence assessment and thus they cannot be directly compared with the Persistence criteria (i.e., REACH R.11 guidance (ECHA, 2023). For example, in OECD (2002) test guideline 308, a water:sediment ratio ($R_{\mathrm{W:S}}$) in the test system of between 3:1 and 4:1 is recommended, but only approximately 20% of studies in the database comply with this recommendation. In addition, important test information may not be reported in nonstandard tests. For example, even though f_{oc} is recognized as a key environmental parameter that impacts partitioning of test substances and therefore biodegradation, approximately 88% of the sediment studies in the database do not report the f_{oc}. Also, nonstandard degradation experiments almost exclusively use nonradiolabeled test compounds which means that nonextractable residues are not quantified or considered in degradation kinetics. This remains a shortcoming in light of updated R.11 REACH guidance (ECHA, 2023).

Much of the data for individual constituents in the database were derived from tests conducted with substances that are UVCBs (i.e., substances of unknown or variable composition, complex reaction products, or biological materials) or that are composed as mixtures. These data differ from those of standard OECD tests, which are typically performed on individual constituents. The HCs generally occur as mixtures in neat petroleum substances that are highly complex UVCBs consisting of a large number of HC constituents. Several concerns arise around the regulatory acceptability of data generated from testing mixtures or multiple constituents at once, including the influence of tested constituents on the degradation of other constituents (ECHA, 2023; Wassenaar & Verbruggen, 2021). However, there is a growing body of supporting literature indicating that testing of mixtures at low, environmentally relevant concentrations does not greatly alter the biodegradation rate and may be more representative (Birch et al., 2022; Hammershøj et al., 2019; Li & McLachlan, 2019; Tian et al., 2023).

The test substance concentration (dose) in the training set is another factor that may affect the accuracy of the predictions. The residual plots provided in Table S5 in the Supporting Information indicate bias at the higher substance concentrations. The objectives of capping the test substance concentration are to avoid inhibitory effects on the microbial population and limit variability in the kinetics (e.g., lag phase) and the extent of biodegradation (OECD, 2006; Strotmann et al., 2023).

Thus the HC-BioSIM training sets (and the subsequent model domain) include a wide range of soil and sediment temperatures, test substance concentrations, organic carbon content, and $R_{\mathrm{w:s}}$. Caution should be exercised when interpolating between values for the environmental or experimental values that are available in the training set. For example, if there is a temperature or dose that is not represented in the training set, the resulting predictions may be less reliable. This caution should also apply to combinations of test system parameters that do not already occur in the training set. It should be noted that the accuracy and reliability of predictions outside the system parameters defined in the training set cannot be fully evaluated at present. Further research is needed to clearly define the HC-BioSIM applicability domain and the reliability of the predictions, as per good modeling standard practice (Buser et al., 2012).

The residual box plots in Table S5 in both the Soil and Sediment Model Supporting Information files provide an indication of the distribution of the data and the range of values for which HC-BioSIM predictions may be more accurate. A more succinct description of the HC-BioSIM applicability domain and test system parameter values based on the most common values in the training set are available as the Supporting Information in the HC-BioSIM QSPR Model Reporting Format (QMRF) document, which was developed according to ECHA guidance (ECHA, 2008).

CONCLUSIONS

The non-standard studies captured in the soil and sediment database were compared with the experimental setup and data interpretation used in the respective OECD simulation test guidelines, through the development of reliability and relevance criteria for soil and sediment biodegradation testing. This allows a more robust database that is weighted accordingly for Persistence assessment in an approach consistent with REACH (ECHA, 2023). The relevance and reliability criteria have been developed to establish representativeness and a means of systematically evaluating non-standard data. Further development and discussion of biodegradation testing methodology are needed to support its application in regulatory assessments.

For a diverse data set of HC constituents and test conditions, HC-BioSIM consistently outperformed the BioHCWin model combined with IMEF for soil and sediment systems. This result is consistent and comparable with previous findings in surface water systems. Additionally, HC-BioSIM provided improved accuracy of Persistence categorization, with correct classification rates of 83.9% and 90.6% for soil and sediment compartments, respectively. Most importantly, however, for the first time, system-specific and environmental effects on the biodegradation of petroleum HCs can be quantitatively evaluated in soil and sediment systems with independent estimation of soil and sediment DT50.

Supporting Information—The Supporting Information is available on the Wiley Online Library at https://doi.org/10.1002/etc.5857.

Acknowledgments—Funding was provided by Concawe (www.concawe.eu). The companion paper to our study can be found at https://doi.org/10.1002/etc.5328.

Author Contributions Statement—**Craig W. Davis:** Conceptualization; Formal analysis; Investigation; Methodology; Validation; Visualization; Writing—original draft. **David M. Brown:** Data curation; Methodology; Validation; Writing—original draft. **Chesney Swansborough:** Data curation; Methodology; Writing—original draft; Writing—review & editing. **Christopher B. Hughes:** Conceptualization; Data curation; Methodology; Supervision; Writing—review & editing. **Louise Camenzuli:** Formal analysis; Methodology; Writing—review & editing. **Leslie J. Saunders**: Conceptualization; Project administration; Supervision; Writing—review & editing. **Delina Y. Lyon**: Project administration; Supervision; Writing—original draft; Writing—review & editing.

Data Availability Statement—The soil and sediment database and all other information are available in the Supporting Information. Additionally, the code for HC-BioSIM is also available on GitHub (https://github.com/concawe/HC-BioSIM-Biodegradation-QSAR-Public).

REFERENCES

Aronson, D., Boethling, R., Howard, P., & Stiteler, W. (2006). Estimating biodegradation half-lives for use in chemical screening. *Chemosphere*, *63*(11), 1953–1960. https://doi.org/10.1016/j.chemosphere.2005.09.044

Aronson, D., Citra, M., Shuler, K., Printup, H., & Howard, P. (1999). *Aerobic biodegradation of organic chemicals in environmental media: A summary of field and laboratory studies*. Prepared for the US Environmental Protection Agency.

Bacosa, H. P., Erdner, D. L., Rosenheim, B. E., Shetty, P., Seitz, K. W., Baker, B. J., & Liu, Z. (2018). Hydrocarbon degradation and response of seafloor sediment bacterial community in the northern Gulf of Mexico to light Louisiana sweet crude oil. *The ISME Journal*, *12*(10), 2532–2543. https://doi.org/10.1038/s41396-018-0190-1

Bai, H., Lu, P., Li, Y., Wang, J., & Zhao, H. (2021). Prediction of phthalate acid esters degradation in soil using QSAR model: A combined consideration of soil properties and quantum chemical parameters. *Ecotoxicology and Environmental Safety*, *226*, 112830. https://doi.org/10.1016/j.ecoenv.2021.112830

Benfenati, E., Manganaro, A., & Gini, G. (2013). *VEGA-QSAR* [Computer software]. www.vegahub.eu

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*(4), 2143–2151. https://doi.org/10.1021/acs.est.7b05624

Birch, H., Sjøholm, K. K., Dechesne, A., Sparham, C., van Egmond, R., & Mayer, P. (2022). Biodegradation kinetics of fragrances, plasticizers, UV filters, and PAHs in a mixture—Changing test concentrations over 5 orders of magnitude. *Environmental Science & Technology*, *56*, 293–301.

Boethling, R. S., Howard, P. H., Beauman, J. A., & Larosch, M. E. (1995). Factors for intermedia extrapolation in biodegradability assessment. *Chemosphere*, *30*(4), 741–752. https://doi.org/10.1016/0045-6535(94)00439-2

Brown, D. M., Camenzuli, L., Redman, A. D., Hughes, C., Wang, N., Vaiopoulou, E., Saunders, D., Villalobos, A., & Linington, 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

Buser, A. M., MacLeod, M., Scheringer, M., Mackay, D., Bonnell, M., Russell, M. H., DePinto, J. V., & Hungerbühler, K. (2012). Good modeling practice guidelines for applying multimedia models in chemical assessments. *Integrated Environmental Assessment and Management*, *8*(4), 703–708. https://doi.org/10.1002/ieam.1299

Davenport, R., Curtis-Jackson, P., Dalkmann, P., Davies, J., Fenner, K., Hand, L., McDonough, K., Ott, A., Ortega-Calvo, J. J., Parsons, J. R., Schäffer, A., Sweetlove, C., Trapp, S., Wang, N., & Redman, A. (2022). Scientific concepts and methods for moving persistence assessments into the 21st century. *Integrated Environmental Assessment and Management*, *18*(6), 1454–1487. https://doi.org/10.1002/ieam.4575

Davis, C. W., Camenzuli, L., & Redman, A. D. (2022). Predicting primary biodegradation of petroleum hydrocarbons in aquatic systems: Integrating system and molecular structure parameters using a novel machine-learning framework. *Environmental Toxicology and Chemistry*, *41*(6), 1359–1369. https://doi.org/10.1002/etc.5328

European Centre for Ecotoxicology and Toxicology of Chemicals (ECETOC). (2009). *ECETOC marine biodegradation kinetics database* (Technical Report No.108 EMBK).

Endo, S., & Goss, K.-U. (2014). Applications of polyparameter linear free energy relationships in environmental chemistry. *Environmental Science & Technology*, *48*(21), 12477–12491. https://doi.org/10.1021/es503369t

Environment and Climate Change Canada. (1999). *Canadian Environmental Protection Act, 1999, S.C. 1999, c.33.* https://laws-lois.justice.gc.ca/eng/acts/c-15.31/page-1.html

European Chemicals Agency. (2008). QSARs and grouping of chemicals. In *Guidance on information requirements and chemical safety assessment.* https://echa.europa.eu/documents/10162/17224/information_requirements_r6_en.pdf/77f49f81-b76d-40ab-8513-4f3a533b6ac9?t=1322594777272

European Chemicals Agency. (2023). PBT/vPvB assessment. In *Guidance on information requirements and chemical safety assessment, Ver. 4.0.* https://echa.europa.eu/documents/10162/17224/information_requirements_r11_en.pdf/a8cce23f-a65a-46d2-ac68-92fee1f9e54f

European Union. (2006). *Regulation (EC) no 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), establishing a European Chemicals Agency, amending Directive 1999/45/EC and repealing Council Regulation (EEC) No 793/93 and Commission Regulation (EC) No 1488/94 as well as Council Directive 76/769/EEC and Commission Directives 91/155/EEC, 93/67/EEC, 93/105/EC and 2000/21/EC.* https://eur-lex.europa.eu/legal-content/en/TXT/HTML/?uri=CELEX:02006R1907-20221217

European Union. (2009). *Regulation (EC) No 1107/2009 of the European Parliament and of the Council of 21 October 2009 concerning the placing of plant protection products on the market and repealing Council Directives 79/117/EEC and 91/414/EEC.* https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:L:2009:309:FULL

GraphPad. (2020). *Prism 9* [Computer software]. https://www.graphpad.com/updates/prism-900-release-notes

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*(6), 3087–3094. https://doi.org/10.1021/acs.est.9b00638

Honti, M., & Fenner, K. (2015). Deriving persistence indicators from regulatory water-sediment studies—Opportunities and limitations in OECD 308 data. *Environmental Science & Technology*, *49*(10), 5879–5886. https://doi.org/10.1021/acs.est.5b00788

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*, *24*(8), 1847. https://doi.org/10.1897/04-453R.1

Li, Z., & McLachlan, M. S. (2019). Biodegradation of chemicals in unspiked surface waters downstream of wastewater treatment plants. *Environmental Science & Technology*, *53*(4), 1884–1892. https://doi.org/10.1021/acs.est.8b05191

Lombardo, A., Manganaro, A., Arning, J., & Benfenati, E. (2022). Development of new QSAR models for water, sediment, and soil half-life. *Science of the Total Environment*, *838*, 156004. https://doi.org/10.1016/j.scitotenv.2022.156004

Organisation for Economic Co-operation and Development. (1992). *Test No. 301: Ready biodegradability*. https://doi.org/10.1787/9789264070349-en

Organisation for Economic Co-operation and Development. (2002). Test No. 308: Aerobic and anaerobic transformation in aquatic sediment systems. *OECD guidelines for the testing of chemicals, Section 3*.

Organisation for Economic Co-operation and Development. (2006). *Revised introduction to the OECD guidelines for testing of chemicals, Section 3*. https://www.oecd-ilibrary.org/content/publication/9789264030213-en

Organisation for Economic Co-operation and Development. (2014). *Guidance document on the validation of (quantitative) structure-activity relationship [(Q)SAR] models*. https://www.oecd-ilibrary.org/content/publication/9789264085442-en

Pavan, M., & Worth, A. P. (2006). *Review of QSAR models for ready biodegradation* (EUR 22355 EN). European Commission Joint Research Centre. https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&cad=rja&uact=8&ved=2ahUKEwjelYPlrumDAxWi4AIHHZRPA_AQFnoECBQQAQ&url=https%3A%2F%2Fpublications.jrc.ec.europa.eu%2Frepository%2Fbitstream%2FJRC34802%2F4802%2520-%2520QSAR%2520Review%2520Biodegradation%255B1%255D.pdf&usg=AOvVaw26drkdxGEDLUuoyQ-iCqSu&opi=89978449

Posada-Baquero, R., Fernández-López, C., Hennecke, D., & Ortega-Calvo, J.-J. (2024). Integrating bioavailability measurements in persistence testing of partially biodegradable organic chemicals in soil. *Science of the Total Environment*, *909*, 168460. https://doi.org/10.1016/j.scitotenv.2023.168460

Quinlan, J. R. (1993). Combining instance-based and model-based learning. *Proceedings of the Tenth International Conference on Machine Learning*, pp. 236–243.

Redman, A. D., Bietz, J., Davis, J. W., Lyon, D., Maloney, E., Ott, A., Otte, J. C., Palais, F., Parsons, J. R., & Wang, N. (2021). Moving persistence assessments into the 21st century: A role for weight-of-evidence and overall persistence. *Integrated Environmental Assessment and Management*, *18*(4), 868–887. https://doi.org/10.1002/ieam.4548

Schäffer, A., Fenner, K., Wang, Z., & Scheringer, M. (2022). To be or not to be degraded: In defense of persistence assessment of chemicals. *Environmental Science: Processes & Impacts*, *24*(8), 1104–1109. https://doi.org/10.1039/D2EM00213B

Seller, C., Özel Duygan, B. D., Honti, M., & Fenner, K. (2021). Biotransformation of chemicals at the water–sediment interface–toward a robust simulation study setup. *ACS Environmental Au*, *1*(1), 46–57. https://doi.org/10.1021/acsenvironau.1c00006

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. https://doi.org/10.1016/j.chemosphere.2019.124516

Shrestha, P., Meisterjahn, B., Klein, M., Mayer, P., Birch, H., Hughes, C. B., & Hennecke, D. (2019). Biodegradation of volatile chemicals in soil: Separating volatilization and degradation in an improved test setup (OECD 307). *Environmental Science & Technology*, *53*(1), 20–28. https://doi.org/10.1021/acs.est.8b05079

Singh, A. K., Bilal, M., Iqbal, H. M. N., & Raj, A. (2021). Trends in predictive biodegradation for sustainable mitigation of environmental pollutants: Recent progress and future outlook. *The Science of the Total Environment*, *770*, 144561. https://doi.org/10.1016/j.scitotenv.2020.144561

Strotmann, U., Thouand, G., Pagga, U., Gartiser, S., & Heipieper, H. J. (2023). Toward the future of OECD/ISO biodegradability testing-new approaches and developments. *Applied Microbiology and Biotechnology*, *107*(7–8), 2073–2095. https://doi.org/10.1007/s00253-023-12406-6

Tian, R., Posselt, M., Fenner, K., & McLachlan, M. S. (2023). Increasing the environmental relevance of biodegradation testing by focusing on initial biodegradation kinetics and employing low-level spiking. *Environmental Science & Technology Letters*, *10*(1), 40–45. https://doi.org/10.1021/acs.estlett.2c00811

United Nations Environment Programme. (2019). *Stockholm convention on Persistent Organic Pollutants (POPS)*. https://chm.pops.int/

US Environmental Protection Agency. (1976). *TSCA as amended by the Frank R. Lautenberg Chemical Safety Act of the 21st Century Toxic substances control* (15 U.S.C. §2601 et seq.). http://uscode.house.gov/view.xhtml?path=/prelim@title15/chapter53&edition=prelim

US Environmental Protection Agency. (2023). *Estimation programs interface Suite™ for Microsoft® Windows, Ver. 4.11* [Computer software].

Wassenaar, P. N. H., & Verbruggen, E. M. J. (2021). Persistence, bioaccumulation and toxicity-assessment of petroleum UVCBs: A case study on alkylated three-ring PAHs. *Chemosphere*, *276*, 130113. https://doi.org/10.1016/j.chemosphere.2021.130113

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*(1), 61. https://doi.org/10.1186/s13321-017-0247-6

Yang, C., Tarkhov, A., Kleinoeder, T., Magdziarz, T., Sacher, O., Schwab, C. H., Schwoebel, J., & Ter, L. (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*(3), 510–528.