

Deriving Persistence Indicators from Regulatory Water-Sediment Studies – Opportunities and Limitations in OECD 308 Data

Mark Honti[†] and Kathrin Fenner^{*,‡,§}

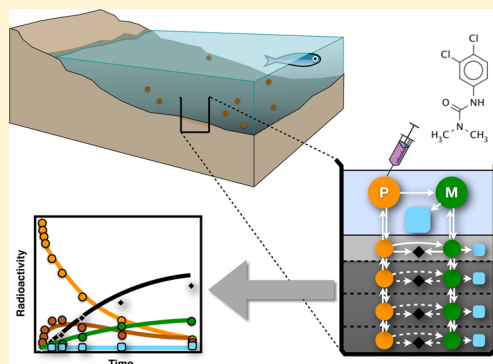
[†]MTA-BME Water Research Group, Hungarian Academy of Sciences, 1111 Budapest, Hungary

[‡]Department of Environmental Systems Science (D-USYS), ETH Zürich, 8092 Zürich, Zürich, Switzerland

[§]Eawag, Swiss Federal Institute of Aquatic Science and Technology, 8600 Dübendorf, Zürich, Switzerland

S Supporting Information

ABSTRACT: The OECD guideline 308 describes a laboratory test method to assess aerobic and anaerobic transformation of organic chemicals in aquatic sediment systems and is an integral part of tiered testing strategies in different legislative frameworks for the environmental risk assessment of chemicals. The results from experiments carried out according to OECD 308 are generally used to derive persistence indicators for hazard assessment or half-lives for exposure assessment. We used Bayesian parameter estimation and system representations of various complexities to systematically assess opportunities and limitations for estimating these indicators from existing data generated according to OECD 308 for 23 pesticides and pharmaceuticals. We found that there is a disparity between the uncertainty and the conceptual robustness of persistence indicators. Disappearance half-lives are directly extractable with limited uncertainty, but they lump degradation and phase transfer information and are not robust against changes in system geometry. Transformation half-lives are less system-specific but require inverse modeling to extract, resulting in considerable uncertainty. Available data were thus insufficient to derive indicators that had both acceptable robustness and uncertainty, which further supports previously voiced concerns about the usability and efficiency of these costly experiments. Despite the limitations of existing data, we suggest the time until 50% of the parent compound has been transformed in the entire system ($\text{DegT}_{50,\text{system}}$) could still be a useful indicator of persistence in the upper, partially aerobic sediment layer in the context of PBT assessment. This should, however, be accompanied by a mandatory reporting or full standardization of the geometry of the experimental system. We recommend transformation half-lives determined by inverse modeling to be used as input parameters into fate models for exposure assessment, if due consideration is given to their uncertainty.



INTRODUCTION

The degradation of chemicals in different environmental compartments is one of the major determinants of their environmental fate and therefore plays a crucial role in regulatory decision-making. On the one hand, degradation half-lives are compared to half-life criteria for persistence assessment in the context of different hazard assessment schemes.¹ On the other hand, degradation rate constants are needed for higher-tier exposure modeling such as, for instance, in the FOCUS models used for the risk assessment of plant protection products and veterinary pharmaceuticals^{2,3} or in the EUSES model used for risk assessment of industrial chemicals and biocides.⁴ Simulation tests, such as those described in the Organisation for Economic Co-operation and Development (OECD) guidelines (OECD 303 for aerobic sewage treatment (activated sludge, biofilms), OECD 307 for soil, OECD 308 for water-sediment systems, OECD 309 for surface waters and OECD 314 for wastewater treatment), are intended to produce the kind of kinetic degradation data that can support modeling and persistence assessment at the higher tiers of regulatory risk assessment. However, there are still a considerable number of open questions

with respect to the performance, evaluation, and interpretation of those studies. Most importantly, the terms persistence and half-life are not consistently defined in different regulatory frameworks, and clear guidance on how they should be derived from simulation study data is lacking.^{5,6}

The OECD 308 guideline on “Aerobic and Anaerobic Transformation in Aquatic Sediment Systems” was adopted in 2002.⁷ It is an integral part of tiered testing strategies in different legislative frameworks for the environmental risk assessment of chemicals. The experiment uses a stagnant and thus fully settled water-sediment system with the parent compound initially spiked into the water phase. Due to the requirement on stagnancy, the system can only be sampled by sacrificial sampling, which then requires up to 30 replicates for one sediment to be run in parallel over a time period of up to 150 days. For regulatory purposes, two such studies, one with a high and one with a low

Received: February 13, 2015

Revised: April 20, 2015

Accepted: April 20, 2015

Published: May 11, 2015

organic carbon content sediment, respectively, are required. The results of these highly laborious and costly studies (approximately 100,000 Euros/compound) are interpreted in various ways, depending on regulatory context. While their primary use in all frameworks is to derive persistence indicators for hazard assessment, or degradation rate constants for exposure assessment/modeling, there are also other endpoints derived from OECD 308 studies. These are the identity, amount, and persistence of transformation products formed, the extent of mineralization, the formation of nonextractable residues, and the extent of mass transferred to the sediment. The last is used as a trigger for sediment ecotoxicity testing of human pharmaceuticals.⁸

Several shortcomings of the OECD 308 guideline have been identified and discussed over the years in workshops,^{9–11} Q&A documents,¹² and the scientific literature.^{13–16} Major points of criticism are related to the lack of environmental relevance (e.g., sediment depth, stagnant conditions, extensive formation of nonextractable residues etc.) and the lack of clarity on how to interpret the study outcomes with respect to the main objective, i.e., to assess a chemical's degradation at the water-sediment interface. The latter difficulties stem from the fact that OECD 308 assesses degradation in a two-phase system where phase transfer processes and degradation happen simultaneously.

In this study, we focus on the latter concern and ask the question of whether and, if so, how robustly degradation information for a given parent chemical can be derived from OECD 308 data. In doing so, our premise is that a robust measure of degradation to be used in chemical risk assessment must only reflect the degradation of a certain chemical given a certain sediment. As such, it should also be reproducible between laboratories, given the same sediment. It should be noted that, throughout this study, we define degradation as the sum of transformation product formation (referred to as metabolites in the following), mineralization, and nonextractable residue formation and do not explicitly address questions of how to interpret nonextractable residue formation. For a more in-depth discussion of this issue, see refs 17–19.

Three types of indicators are typically derived from OECD 308 data. $DT_{50,w}$ and $DT_{50,sed}$ are the time until 50% of the parent chemical has disappeared from the water phase or from the sediment phase after reaching its peak concentration. These indicators are easily accessible since they are directly observable from the data and can be quantified using interpolation between measured time points. However, both of these indicators clearly lump together phase transfer and degradation processes and therefore cannot be considered as robust degradation measures in the above-defined sense.¹⁶

In contrast, $DegT_{50,system}$, which is the time until 50% of the parent chemical has disappeared from the system and which is also directly observable from the data,¹⁵ is clearly a degradation indicator since the parent chemical can only disappear from the total system through transformation to either metabolites or nonextractable residues. Thus, $DegT_{50,system}$ is a potential candidate for an easily accessible indicator of persistence at the water-sediment interface. However, it should certainly not be used as half-life for exposure modeling where different environmental compartments are typically treated separately and hence compartment-specific degradation rate constants are required.

Therefore, $DegT_{50,w}$ and $DegT_{50,sed}$, which are the actual degradation half-lives of the chemical in the water and sediment compartment, respectively, have been suggested as measures of

degradation that are fully independent of phase transfer processes and are potentially suitable for both persistence assessment and exposure modeling.²⁰ However, they are not directly observable from the data and therefore have to be derived by inverse modeling. Such decomposition of the measurement data necessarily introduces considerable uncertainty into the estimated kinetic parameters due to model uncertainty and identifiability issues, i.e., phase transfer and degradation rate constants can compensate for each other, thus yielding similarly good fits to the data for a wide range of parameter values that hence are difficult to estimate. Due to the complexity of simulating the OECD 308 system, inverse modeling has been rarely attempted. Radke et al. (2009)²¹ carried out the estimation of transformation rate constants on a variant of OECD 308 but neither discussed model uncertainty nor assessed corollary parameter uncertainty.

Altogether, there is an obvious trade-off between conceptual robustness and uncertainty in the derivation of degradation measures from OECD 308 data. The objective of this paper therefore is to make this dilemma transparent and quantitative such as to deliver a sound basis for deciding which indicator is best used in which regulatory context and what the “best practice” should be for using this indicator. To reach this goal, we (i) develop an inverse modeling framework to quantify $DegT_{50,w}$ and $DegT_{50,sed}$ values and their uncertainties for different model hypotheses on the processes taking place in the water-sediment system and apply this framework to 23 pesticides and pharmaceuticals spanning a wide spectrum of degradation and sorption behavior; (ii) explore the uncertainties of $DegT_{50,system}$ as a function of different experimental systems, and (iii) based on the results of (i) and (ii) give recommendations on the suitability and best-practice of using $DegT_{50,system}$ and $DegT_{50,w}$ and $DegT_{50,sed}$, respectively, as measures for persistence assessment and exposure modeling.

METHODS

Data. We analyzed 41 experimental OECD 308 data sets for 23 different pharmaceuticals ($n = 7$) and pesticides ($n = 16$). Compounds were selected based on data availability and to adequately cover the range of sorption behavior (i.e., organic carbon–water partition coefficients) and persistence (i.e., $DegT_{50,system}$ values from OECD 308 testing) represented in the Pesticide Properties DataBase (<http://sitem.herts.ac.uk/aeru/ppdb/>). For pesticides, the OECD 308 data were read from Draft Assessment Reports (DARs) made publicly available by the European Food Safety Authority.²² For the seven pharmaceuticals, including three anonymous compounds, the data were provided by two pharmaceutical companies. For the majority of compounds (18 out of 23) data from two experiments were available, one carried out with a sediment possessing a higher and one with a lower organic matter content as required by the 308 guideline. A list of compounds and available data sets is given in Table S1 in the Supporting Information (SI).

Model Structure. The stagnant water-sediment setup in OECD 308 experiments leads to strong heterogeneity within the small experimental vessels. The lack of stirring means that diffusion governs all transport processes, which may result in concentration gradients along the 2–3 cm deep sediment layer. Specifically, we experimentally confirmed that, except for a very shallow surface layer, the sediment is anoxic because of O_2 consumption taking place in the sediment and the limited supply through the stagnant water body (see Figure S1 in the SI). Besides this, there is a slowly developing concentration gradient

of the parent compound and the metabolites as they penetrate deeper into the sediment via diffusion in porewater. This diffusion is slowed-down to a compound-specific degree by sorption to the immobile sediment particles. As the above-mentioned gradients are packed into a shallow vertical space, small changes in the geometry of the actual experimental systems may exert a considerable influence on the compounds' behavior. Despite the obvious importance of these boundary conditions, the OECD 308 guideline does not specify strict rules for geometrical parameters (sediment depth can vary between 2 and 3 cm, water:sediment volume ratio between 3 and 4). While a proper mathematical model of the OECD 308 system thus has to account for the heterogeneities in the sediment to produce a good fit to measured data, no complementary data on concentration gradients are typically measured in OECD 308 studies.

We therefore created four different but closely related mechanistic models that represent the same OECD 308 experimental system with different hypotheses on compound fate in the sediment. Several assumptions applied to all versions, as follows. All models used two compartments: chemicals inside the system were assumed to be either in dissolved phase in the water column (i.e., neglecting sorption to dissolved organic carbon) or in sorbed or dissolved state in the sediment compartment. All transformations were assumed to follow first-order kinetics in both compartments. Diffusion into the sediment was described by Fick's law. Sorption equilibrium in the sediment was assumed to be reached instantaneously. Transformation pathways were identical in all versions. The parent compound initially added to the water compartment could be either transported to the sediment or transformed to undifferentiated metabolites in the water column. In the sediment, the parent compound could be transformed either to metabolites or nonextractable residues (NER) and metabolites could be transformed to CO₂ or NER.

The four model versions differ in the chemical and physical fate of the compounds in the sediment compartment (see Figure S2 in the SI). In model A, we assumed instantaneous mixing along the entire depth of the sediment for both parent compound and metabolite(s) when entering the sediment column. In this case, the sediment compartment itself was treated as a fully mixed reactor, i.e., transformation processes were assumed to take place synchronously and to the same extent throughout the entire sediment. This is in line with recommendations given in the FOCUS Guidance Document on Estimating Persistence and Degradation Kinetics from Environmental Fate Studies²⁰ at level P-II. In model B, we accounted for slow diffusion into the sediment, calculating an effective sediment depth that increases with time as suggested in Appendix IX of the FOCUS Guidance Document.²⁰ Transformation was thus limited to the actual penetration depth. As the sediment is typically anaerobic except for the thin surface layer (see Figure S1 in the SI), this implicitly assumes equal aerobic and anaerobic transformation rate constants. Therefore, in versions C and D, diffusion into the sediment was explicitly simulated by splitting the sediment column into 4 horizontal compartments. In version C, transformation was assumed to be exclusively aerobic and to take place in the uppermost boundary layer only. In version D, transformation was allowed in the deeper anoxic layers too, albeit at a reduced rate. More details on model versions A to D and all model equations are given in the SI.

Parameter Inference. Although our 4 model versions are still simplified descriptions of the true system, they already

possess many degrees of freedom with regard to calibration. Version D has 15 model parameters (including 6 transformation rate constants), 6 initial values, and 1 error parameter. Versions A, B, and C have 14 model parameters, 6 initial values, and again 1 error parameter. These relatively large parameter sets would be difficult to calibrate based on experimental data alone. The measurements follow certain compound pools in the system, while parameters govern processes that move the compounds between the measured pools. Due to the complexity of the models several processes can compensate for each other's effect, which leads to identifiability problems. These in turn lead to infinite sets of totally different model parameter values producing a similarly good fit to the data.

We used Bayesian parameter inference to overcome the identifiability problems in a transparent and formal statistical way. Bayesian parameter inference has already been used in modeling transformation kinetics^{18,23} but exclusively with relatively simple, single-compartment models and wide, substance-independent priors. Here, existing information that is independent of the measurements themselves was formulated as prior probability distributions for the parameters (Table S3 in the SI). Organic carbon–water partition coefficients (K_{oc}) determined in different soils were used to define priors for K_{oc} of the parent compound and for the ratio of K_{oc} of the parent and the most relevant metabolite(s) (Δ). Experiment-specific data were used for the organic content of the sediment. For the remaining parameters, we used expert estimations where available (e.g., diffusion coefficient was set to mean = 0.5 and standard deviation = 0.5 cm² d⁻¹ based on ref 24) and broad general priors according to our expectations where there was no prior information whatsoever.

Parameters were then inferred following Bayes' rule: the posterior probability of a certain parameter set was proportional to the prior probability of the set times the likelihood of the set given the experimental data (see, for instance, ref 23). The likelihood was calculated with a simple frequentist error model that assumed independent and identically distributed normal errors. The standard deviation of the error was calibrated together with other model parameters.²³ While it would have been possible to only look for the maximum posterior solution that best fits the data, we considered the uncertainty of our parameter estimates at least as important as the best values. To discover the parameter posteriors we used Markov Chain Monte Carlo (MCMC) sampling using the traditional Metropolis algorithm.²⁵ The covariance matrix of the proposal distribution was tuned during the burn-in phase to keep acceptance between 15 and 40%.²⁶ Three MCMC chains were generated to ensure convergence and stability, each having a total length of 200,000 rounds with 100,000 rounds dedicated for burn-in. Samples were thinned by a factor of 5 to reduce serial correlation. The Bayesian parameter inference tool in combination with all four model versions is available as downloadable software (<http://www.eawag.ch/en/departement/uchem/software/>). It can be used to determine the posterior densities of transformation rate constants from OECD 308 data as described in this paper.

Analyses of Parameter Robustness. Several sensitivity analyses were carried out to discover the robustness of various persistence indicators (Table 1). The necessity of the Bayesian parameter inference procedure (involving independent information in addition to the experimental data in the form of informative prior distributions) was checked by attempting a parameter uncertainty analysis with fully vague priors for each parameter (uniform distribution on the [0,∞[interval or over

Table 1. Summary of Comparative Analyses Targeted at Parameter Robustness

model version	priors	sediments	target of analysis
D	vague vs standard vs narrow	individual	necessity of Bayesian calibration and influence of prior distributions
A vs B/C/D	standard	individual	diffusion in sediment
B vs C	standard	individual	aerobic vs anaerobic transformation in sediment
D	standard	individual vs paired	universality of transformation rate constants among sediments

the physically meaningful range) for model version D. The outcome was compared to the results of version D with the standard priors. The sensitivity of transformation rate constants to the (informative) prior distributions was tested by comparing the output of model version D (the most flexible) using the standard and a stricter set of prior distributions for K_{oc} of the parent and metabolites and diffusion coefficients. The importance of describing diffusion in the sediment was tested by comparing the quality of fit in model version A (assumes mixed sediment) vs versions B, C, and D (assume gradual diffusion into sediment). The role of hypotheses about the place of transformation in the sediment was tested by comparing transformation rate constants derived from model versions B and C, because these model versions formulated the most extreme hypotheses about the relative importance of aerobic and anaerobic degradation. Finally, we calibrated model version D on individual experiments alone and also in combination for compounds possessing two data sets from different sediments. In this way, we could test whether the experimental data provided any evidence to refute the universality hypothesis (i.e., that the transformation rate constant was the same in both sediments) and, thus, indirectly test the validity of our assumption that transformation rate constants differ across sediments for an individual compound.

Analysis on the Influence of System Geometry on $\text{DegT}_{50, \text{system}}$. The OECD 308 guideline criteria regarding the geometry of the experimental systems are somewhat flexible. Sediment depth must be between 2 and 3 cm, and the water:sediment volumetric ratio must be between 3 and 4. This can ultimately result in water column heights varying between 6 and 12 cm. These varying system geometries can influence apparent persistence in the experimental system. For instance, for compounds that are transformed faster in the sediment than in the water column, the system can theoretically be manipulated to show a relatively shorter $\text{DegT}_{50, \text{system}}$ by using a lower water:sediment ratio. To investigate the possible magnitude of this influence we ran a series of numerical experiments under the smallest and largest allowed water:sediment ratios with model versions B, C, and D and the maximum posterior likelihood parameters of our experimental compounds. $\text{DegT}_{50, \text{system}}$ was calculated as disappearance half-life of the parent compound using the single first order (SFO) method suggested in the FOCUS guideline.²⁰

RESULTS AND DISCUSSION

Necessity of Bayesian Parameter Inference and Robustness against Priors. The power of Bayesian parameter inference depends on the prior distributions. Using vague priors diminishes the advantage of incorporating existing knowledge and weakens the ability to resolve identifiability problems.

Using very confident (narrow) priors brings the danger of relying strongly on potentially wrong information, which would introduce a significant bias in the parameters.

Model version D with vague priors could achieve equally good or even better fits to the experimental data than the same model coupled with the standard priors. This was expected because Bayesian parameter inference with informative priors actually forces a compromise between model fit and compliance to the expectations about the parameters. Thus, from a strictly mathematical point of view, calibrating a process-based model on OECD 308 data requires neither informative priors nor a Bayesian calibration procedure. However, due to the high number of interacting model parameters, it is not guaranteed that the calibrated parameter values conform to the modeler's knowledge or expectations. Compared to the standard prior set, vague priors introduced convergence problems into MCMC sampling and increased the uncertainty (in terms of bias and variability) of estimated parameters. Inferred parameters often violated physical constraints, expectations based on common sense, and recommended OECD 308 system parameters. Posterior transformation rates were more robustly estimated but sometimes reached extreme values (up to 10^4 d^{-1} ; Figure S3 in the SI).

The comparison of the standard and narrow priors sets revealed that transformation rate constants were in general very weakly influenced by the informative prior distribution sets (Figure S4 in the SI). There were three exceptional data sets where the narrow priors prevented achieving good fit, probably due to the very confident but possibly wrong priors. The other 37 experiments showed robustness against the adjustment of informative priors, which indicated that the data sets contain some relatively robust information on transformation rate constants.

Robustness against Hypotheses on Diffusion Inside Sediment. The comparison of maximal posterior probabilities for model versions A versus B, C, and D revealed that model version A failed to attain a good fit to the data in many more cases than the more complicated model versions B–D (Figure S5 in the SI). The total log posterior probability was significantly lower for the entire data set for model version A than for the others (Figure S6 in the SI). Model version D was the most successful due to its highest structural flexibility. This comparison proved that the simulation of the gradual penetration into the sediment was a necessary feature to describe the experimental data with high fidelity. This finding is actually in line with the recommendations of Annex IX of the FOCUS guidance document to consider the effective depth a substance has moved into the sediment when calculating the expected fraction in the sediment to restrain parameter estimation.²⁰ However, when reviewing the evaluations of OECD 308 data presented in the pesticide DARs, we found that these recommendations are not implemented in practice.

Robustness against Hypotheses on Transformation Inside Sediment. Version A was proven to be a true outlier in terms of quality of fit because it does not account for diffusion in the sediment. The other versions B, C, and D achieved roughly similar log posterior probability scores despite their different assumptions on transformation mechanism. As version D was capable of behaving like either versions B and C, it provided the best fit for most compounds. However, the quality gain by its additional parameter (∂ , which provides version D with the ability to behave like B [$\partial=1$] or C [$\partial=0$]) was not enough to yield a much better, i.e., lower, Akaike Information Criterion

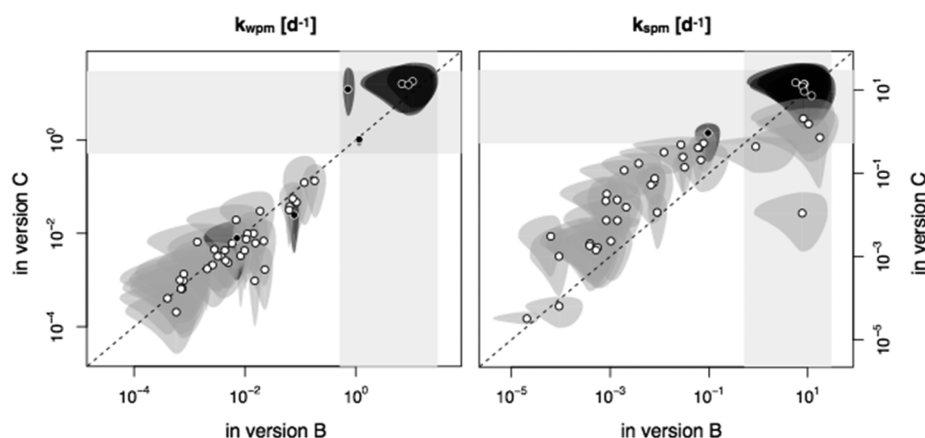


Figure 1. Transformation rate constants of the parent compound in water (k_{wpm}) and in sediment (k_{spm}) in model version B versus version C. The gray bands indicate the 90% confidence interval of the prior distribution. Black points indicate compounds that are believed to hydrolyze quickly. The amorphous regions show the 95% posterior uncertainty interval.

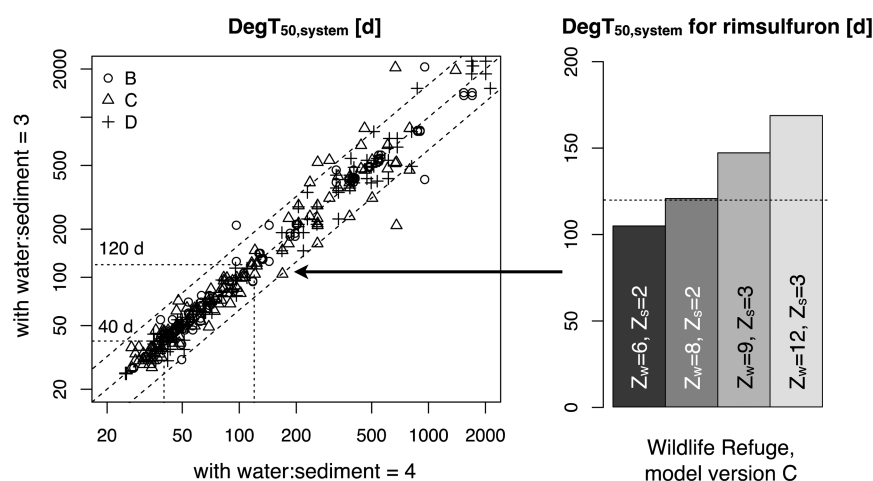


Figure 2. Influence of experimental system geometry on $\text{DegT}_{50,\text{system}}$ with different model versions (left). Dashed lines show the standard deviation of all relative differences (60%). The experiment with rimsulfuron @ Wildlife Refuge with model version C is shown to illustrate a case where the allowed variation in system geometry may push $\text{DegT}_{50,\text{system}}$ above or below a hypothetical persistence cutoff value of 120 days (right). Z_w and Z_s are given in [cm].

(AIC)²⁷ score (B: 11756; C: 11533; D: 11245). This indicated that the experimental data did not provide any evidence about where transformation was actually taking place along the sediment profile. Thus, there was not enough information to decide between the mutually exclusive hypotheses of version B and C or the compromise hypothesis manifested in version D.

The similarly good performance by entirely different model versions still had severe consequences on transformation rate constants (Figure 1). The assumption in C relative to B was that transformation in the sediment took place in a thin surface layer, which enclosed only a small fraction of the total sediment pool of the compound once it had started to diffuse into the sediment. To produce the observed amount of metabolites, CO_2 , and NER, sediment transformation rate constants were estimated to be up to an order of magnitude higher in version C than in version B. Values for version D were in between the two extremes. In contrast, since the water compartments of the four model versions were the same, transformation rate constants in water were hardly affected by differences in model versions B, C, and D. In Table S4 in the SI the estimated parent transformation rate constants for model version A (FOCUS default model) and model version D (best fitting model) are given.

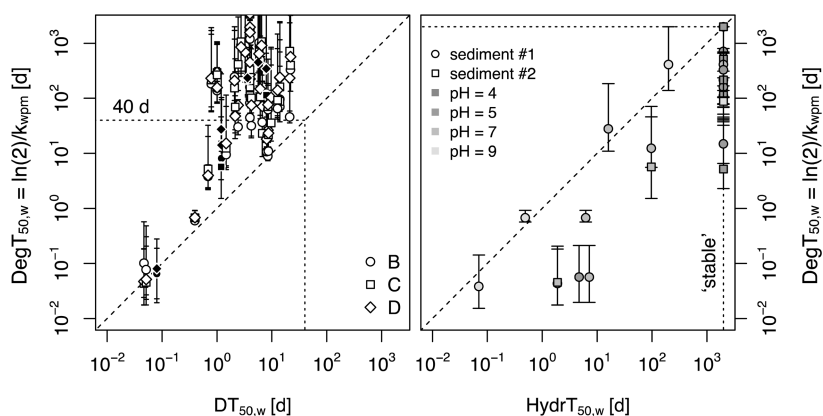
Universality of Transformation Rate Constants among Sediments.

Parameter inference on data sets from individual experiments makes the implicit assumption that transformation rate constants may differ across sediments for the same compound. This is in line with the concept that transformation is strongly dependent on sediment type. The calibration to individual experiments apparently supported this hypothesis because typically different transformation rate constants were inferred from sediment pairs belonging to the same compound. Alternatively, one can introduce an expectation that transformation rate constants are universal among sediments by performing a joint calibration of data sets belonging to the same compound but different sediments.

Interestingly, although joint calibration could certainly not surpass the quality of fit from the solutions that considered a single experiment alone, the fit was not significantly worse than the fits obtained with the individual experiments in most cases (12 out of 15 compounds had less than 25% difference in the calibrated standard deviation of model error). At the same time, universal transformation rate constants differed significantly for both water and sediment ($p < 0.013$ in all cases) from the results of individual experiments for several compounds (Figure S7 in

Table 2. Comparative Analysis of Different Persistence Indicators Derivable from OECD 308 Data with Respect to Conceptual Robustness and Uncertainty

	indicators		
	DegT _{50,system}	DegT _{50,w}	DegT _{50,sed}
conceptual robustness	– not fully robust, depends on system geometry	+ robust, not confounded by phase transfer	
uncertainty	+ little uncertainty, directly observable from data	– uncertain, has to be inferred by modeling; consistent across model options	– highly uncertain, has to be inferred by modeling; not consistent across model options/assumptions
suitability for risk assessment	persistence assessment only	persistence assessment and exposure modeling	
conclusions	easy to observe but easy to use inappropriately	promising because of conceptual robustness but currently highly uncertain	

**Figure 3.** Comparison of different half-lives of the parent compound in water. Left: Estimated transformation half-lives in water (DegT_{50,w}) versus DT_{50,w} with model versions B, C, and D. Right: Estimated transformation half-lives in water with model version D versus chemical hydrolysis half-lives in water (HydrT_{50,w}) reported in EFSA DAR documents. Whiskers show 95% posterior uncertainty intervals in both panels. Compounds found to be stable with respect to chemical hydrolysis are shown at HydrT_{50,w} = 2000 d.

the SI). This suggests that there is no clear evidence in the data about transformation rate constants being sediment-specific, because the hypothesis on the universality of the transformation rate constants is performance-wise equivalent to the assumption of sediment-specific transformation rate constants.

Robustness against Variations in System Geometry.

The allowed variation in the water:sediment ratio can indeed influence DegT_{50,system} (Figure 2). There was up to 40% relative difference in DegT_{50,system} between the “thin” (water:sediment = 4) and “thick” (water:sediment = 3) sediment systems until 500 days of DegT_{50,system}, which is the maximally relevant range considering the typical duration of OECD 308 experiments. Afterward the difference grew, sometimes exceeding 100%. Interestingly, neither the sign of difference, nor the model version in which the difference showed up followed a clear pattern, but DegT_{50,system} was the most extreme for the Z_w = 6 cm, Z_s = 2 cm and Z_w = 12 cm, Z_s = 3 cm pairs in all cases. This suggests that without knowing anything about actual system geometry, DegT_{50,system} may, on average, have at least 40% a priori uncertainty due to the sloppily defined experimental geometry standards. This can influence the classification of compounds as (non)persistent: in our data set of 41 experiments and 3 model versions (B,C,D), DegT_{50,system} crossed the threshold value for persistence in water of 40 days for 7 cases and the threshold values for persistence in sediment of 120 days for 2 cases, within the allowed variation of system geometry (Figure 2).

CONSEQUENCES FOR USING OECD 308 DATA IN A REGULATORY CONTEXT

There is a disparity between the uncertainty and the environmental relevance of persistence indicators derivable from OECD 308 data (Table 2). Indicators such as DegT_{50,system}, DT_{50,w} and DT_{50,sed} are directly extractable from the data with limited uncertainty, but they are specific to the experimental system. Less system-specific indicators (such as DegT_{50,w} and DegT_{50,sed}) require a rather involved inverse modeling approach to extract and suffer from considerable uncertainty.

The simplest and least uncertain indicator, the disappearance time of the parent compound in the total system (DegT_{50,system}), depends on system geometry and as such does not fully qualify as a robust measure of persistence. Based on our analysis, the degrees of freedom in system geometry that are given in the current OECD 308 guideline result in about 40% uncertainty in DegT_{50,system} in the relevant half-life range.

As pointed out in the Introduction, phase-specific apparent persistence indicators (DT_{50,w} and DT_{50,sed}) cannot be considered as robust degradation measures because they lump together transformation and phase transfer. This is confirmed by the comparison of DegT_{50,w} values estimated in this study with DT_{50,w} values derived from the same data as presented in Figure 3. DegT_{50,w} was found to be always much longer than DT_{50,w}. This indicated that DT_{50,w} was actually dominated by diffusion, and transformation only played a secondary role. At the same time, DegT_{50,w} was always smaller than or equal to chemical hydrolysis half-lives reported in the EFSA DAR documents,²² which confirmed that transformation in the water column of OECD 308 studies also reflects processes other than chemical

hydrolysis, most likely pelagic biotransformation (Figure 3). It is of further interest to note that using $\text{DegT}_{50,w}$ instead of $\text{DT}_{50,w}$ for comparison against the persistence threshold for freshwater (40 d), as suggested in the REACH draft guidance for PBT assessment,²⁸ results in a promotion from nonpersistent to persistent for the majority of substances investigated (Figure 3).

Less system-specific indicators that exclude phase transfer ($\text{DegT}_{50,w}$ and $\text{DegT}_{50,\text{sed}}$) are conceptually more robust, but this advantage comes at the cost of elevated uncertainty. According to our results, $\text{DegT}_{50,w}$ was identifiable and similar for all models, most probably due to the identical description of the water phase in all model versions. Even so, however, the estimated $\text{DegT}_{50,w}$ values were still subject to large parametric uncertainty (CVs in the range of 0.5–3.0) resulting from inverse modeling. On top of this parametric uncertainty, $\text{DegT}_{50,\text{sed}}$ showed a strong dependence on the assumptions on transformation mechanisms in the sediment. Available data did not provide a way to decide between the competing assumptions on the compounds' fate in the sediment, so the different (uncertain) estimates would need to be merged in an even more uncertain composite estimate to get a real picture about the uncertainty of $\text{DegT}_{50,\text{sed}}$. Moreover, the universality of $\text{DegT}_{50,\text{sed}}$ values among different sediments could neither be proven nor refuted. The competing hypotheses of $\text{DegT}_{50,\text{sed}}$ values being sediment-specific or universal for a certain compound delivered statistically equivalent fits to the data. However, the two hypotheses resulted in significantly different estimates, which pointed toward a so far not quantified uncertainty.

Despite our sophisticated inverse modeling approach, the data available from experiments carried out according to the OECD 308 guideline were thus insufficient to derive persistence indicators that had both acceptable levels of robustness and uncertainty at the same time. Based on these findings, we would like to outline two strategies: (i) a strategy to extract the maximal, yet defensible amount of persistence information from existing OECD 308 data, and (ii) a strategy to improve our ability to assess persistence at the sediment-water interface in future testing.

With regard to using existing OECD 308 data, we suggest that $\text{DegT}_{50,\text{system}}$ values can be used as indicators of persistence at the water-sediment interface in the context of PBT assessment and for the purpose of benchmarking compounds against each other. It should, however, be noted that this persistence estimate is most representative of degradation in the uppermost, at least partially aerobic sediment layer in a small, shallow, stagnant, or slowly flowing water body. To improve the conceptual robustness of using $\text{DegT}_{50,\text{system}}$ values as persistence indicators, we recommend to either revise the existing OECD 308 guideline to further restrict the allowed variability in system geometry or to strictly enforce reporting of the applied system geometry (i.e., the test vessel's inner diameter, the heights of the water and sediment column, and sediment dry weight used). Similar considerations might also be warranted for other instances of experimental freedom in the current guideline, e.g., different approaches for aeration of the system. One clear advantage of $\text{DegT}_{50,\text{system}}$ is that it could be derived from the amount of extractable parent in the total system and hence would not require the separation of the sediment and water phase, which is hard to standardize. However, the opportunities for doing so might be limited in practice since most regulatory frameworks require other data to be extracted from OECD 308 studies that do require phase separation. Finally, proper usage of $\text{DegT}_{50,\text{system}}$ values derived from OECD 308 data as persistence indicator for PBT

assessment would require the definition of a persistence threshold against which these values should be benchmarked.

For environmental exposure assessment purposes, neither $\text{DT}_{50,w}$ nor $\text{DT}_{50,\text{sed}}$ should be used, but rather inverse modeling as presented here should be used to derive $\text{DegT}_{50,w}$ and $\text{DegT}_{50,\text{sed}}$ values. These can then defensibly be used as input parameters into models for environmental fate and exposure assessment. However, we advise that their uncertainty should be considered quantitatively in those models, for instance by means of Monte Carlo simulations.²⁹ While we are confident that the uncertainty of $\text{DegT}_{50,w}$ obtained from inverse modeling adequately presents the true uncertainty of this parameter, the uncertainty estimates obtained for $\text{DegT}_{50,\text{sed}}$ most likely underestimate its true uncertainty due to the underlying model uncertainties. Even when accounting for these uncertainties, however, the Bayesian approach still leads to a clear gain in information when compared to the state-of-the-art presented in the FOCUS Guidance Document on Estimating Persistence and Degradation Kinetics from Environmental Fate Studies.²⁰ There it is suggested to use default worst-case half-lives of 1000 d for exposure modeling if fitting of the 2-compartment model fails. Our results indicate that for >90% of the substance/sediment combinations evaluated, resulting half-lives in both water and sediment were lower than 1000 d.

It is a common critique against the OECD 308 experimental setup that it has limited direct relevance for water bodies having higher water:sediment ratios, such as most real streams and lakes. Nevertheless, transformation rates derived from 308 experiments could be theoretically used to estimate $\text{DegT}_{50,\text{system}}$ for these different exposure situations. When doing so, the results show that such estimated "environmental" $\text{DegT}_{50,\text{system}}$ values vary largely for a range of realistic water:sediment ratios (Figure S8 in the SI). For almost all compounds, there are specific water:sediment ratios that either make the compound lie above or below typical persistence threshold values.

Given the low amount of high quality persistence information actually extractable from OECD 308 data and the considerable costs involved in carrying out such studies, it seems imperative to think about possibilities to enhance their information content. Our analysis shows that involvement of additional data (e.g., well-specified system geometries, sediment redox conditions) would help to reduce the uncertainty of persistence indicators derived from OECD 308 experiments. Also, a combined evaluation of OECD 308 data with data from stirred systems (e.g., tests carried out according to the OECD 309 guideline) could theoretically help reduce uncertainty but first needs to be tested in practice. Nevertheless, the inherent complexity of the heterogeneous experimental system will most probably continue to cause problems in data analysis. Therefore, stirred tests with suspended sediments along the lines suggested in the OECD 309 guideline could turn out to be cost-efficient, easy to evaluate replacements or complements for assessing degradation at the sediment-water interface.

■ ASSOCIATED CONTENT

● Supporting Information

Data on the compounds and sediments used in this study, a more detailed description of the model versions and various materials on the outcome of parameter inference experiments. The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.est.5b00788.

■ AUTHOR INFORMATION

Corresponding Author

*Phone: +41 58 765 5085. E-mail: kathrin.fenner@eawag.ch,

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors acknowledge Rani Bakkour for extracting and electronically encoding OECD 308 data from EFSA DARs; Pfizer Inc. and Bayer AG for providing OECD 308 data for pharmaceuticals; Stefan Hahn, Dieter Hennecke, Thomas Junker, Martin Scheringer, Johannes Ranke, Silvia Berkner, and the members of the ECETOC research liaison team for critically reading a previous version of the manuscript. This study is funded by the CEFIC Long-range initiative under project identifier LRI-ECO18-Eawag.

■ REFERENCES

- (1) Boethling, R.; Fenner, K.; Howard, P.; Klecka, G.; Madsen, T.; Snape, J. R.; Whelan, M. J. Environmental Persistence of Organic Pollutants: Guidance for Development and Review of POP Risk Profiles. *Int. Environ. Assess. Manage.* **2009**, *5*, 539–556.
- (2) *Surface Water Models and EU Registration of Plant Protection Products*; Report of the work of the Regulatory Modelling Working Group on Surface Water Models of FOCUS (FORum for the Coordination of pesticide fate models and their USE); 1997.
- (3) *Revised Guideline on Environmental Impact Assessment for Veterinary Medicinal Products in Support of the VICH Guidelines GL6 and GL38*; EMEA/CVMP/ERA/418282/2005-Rev.1; Committee for Medicinal Products for Veterinary Use (CVMP); European Medicines Agency (EMA): London, 2009.
- (4) *Guidance on information requirements and chemical safety assessment, Chapter R.16: Environmental Exposure Estimation*; European Chemicals Agency (ECHA): Helsinki, 2008.
- (5) Rauer, C.; Friesen, A.; Hermann, G.; Johncke, U.; Kehr, A.; Neumann, M.; Prutz, I.; Schonfeld, J.; Wiemann, A.; Willhaus, K.; Woltjen, J.; Duquesne, S. Proposal for a harmonised PBT identification across different regulatory frameworks. *Environ. Sci. Eur.* **2014**, *26*, 9.
- (6) Solomon, K.; Matthies, M.; Vighi, M. Assessment of PBTs in the European Union: a critical assessment of the proposed evaluation scheme with reference to plant protection products. *Environ. Sci. Eur.* **2013**, *25*, 10.
- (7) *OECD guidelines for testing of chemicals; 308, Aerobic and anaerobic transformation in aquatic sediment systems*; Organisation for Economic Co-operation and Development (OECD): Paris, 2002.
- (8) *Guideline on the Environmental Risk Assessment of Medicinal Products for Human Use*; Committee for Medicinal Products for Human Use (CHMP); European Medicines Agency (EMA): London, 2006.
- (9) Bowmer, T.; Leopold, A. *Strategies for selecting biodegradation simulation tests and their interpretation in persistence evaluation and risk assessment; Simulation Testing of Environmental Persistence (STEP)*; Report of a two-day workshop held in Rotterdam, 4 and 5 October, 2004; Rotterdam, 2004.
- (10) *Workshop Report No. 17: Significance of bound residues in environmental risk assessment, 14–15 October 2009, Brussels*; European Centre for Ecotoxicology and Toxicology of Chemicals (ECETOC): Brussels, 2010.
- (11) *Special Report No. 18: Information to be considered in a weight-of-evidence-based PBT/vPvB assessment of chemicals*; European Centre for Ecotoxicology and Toxicology of Chemicals (ECETOC): Brussels, 2014.
- (12) *Questions and answers on 'Guideline on the environmental risk assessment of medicinal products for human use'*; European Medicines Agency, Committee for Medicinal Products for Human Use (CHMP): London, 2011.
- (13) Davis, J. W.; Gonsior, S.; Marty, G.; Ariano, J. The transformation of hexabromocyclododecane in aerobic and anaerobic soils and aquatic sediments. *Water Res.* **2005**, *39*, 1075–1084.
- (14) Ericson, J. F. An evaluation of the OECD 308 water/sediment systems for investigating the biodegradation of pharmaceuticals. *Environ. Sci. Technol.* **2007**, *41*, 5803–5811.
- (15) Ericson, J. F.; Smith, R. M.; Roberts, G.; Hannah, B.; Hoeger, B.; Ryan, J. Experiences with the OECD 308 transformation test: A human pharmaceutical perspective. *Int. Environ. Assess. Manage.* **2014**, *10*, 114–124.
- (16) Radke, M.; Maier, M. P. Lessons learned from water/sediment testing of pharmaceuticals. *Water Res.* **2014**, *55*, 63–73.
- (17) Kästner, M.; Nowak, K. M.; Miltner, A.; Trapp, S.; Schäffer, A. Classification and modelling of non-extractable residue (NER) formation of xenobiotics in soil—a synthesis. *Crit. Rev. Environ. Sci. Technol.* **2014**, *44*, 2107–2171.
- (18) Loos, M.; Krauss, M.; Fenner, K. Non-extractable residue formation: Insights from kinetic meta-analysis of regulatory soil simulation studies. *Environ. Sci. Technol.* **2012**, *46*, 9830–9837.
- (19) Matthies, M.; Witt, J.; Klasmeier, J. Determination of soil biodegradation half-lives from simulation testing under aerobic laboratory conditions: A kinetic model approach. *Environ. Pollut.* **2008**, *156*, 99–105.
- (20) *Guidance Document in Estimating Persistence and Degradation Kinetics from Environmental Fate Studies in Pesticides in EU Registration*; The Final Report of the Work Group on Degradation Kinetics of FOCUS (FORum for the Coordination of pesticide fate models and their USE), EC Document Reference Sanco/10058/2005, version 2.0; 2006.
- (21) Radke, M.; Lauwigi, C.; Heinkele, G.; Murdter, T. E.; Letzel, M. Fate of the antibiotic sulfamethoxazole and its two major human metabolites in a water sediment test. *Environ. Sci. Technol.* **2009**, *43*, 3135–3141.
- (22) *Rapporteur member state assessment reports submitted for the EU peer review of activesubstances used in plant protection products*; European Food Safety Authority (EFSA), 2014. <http://dar.efsa.europa.eu/dar-web/provision>.
- (23) Grolitz, L.; Gao, Z. L.; Schmitt, W. Statistical analysis of chemical transformation kinetics using Markov-Chain Monte Carlo methods. *Environ. Sci. Technol.* **2011**, *45*, 4429–4437.
- (24) Schwarzenbach, R. P.; Imboden, D. M.; Gschwend, P. M. *Environmental Organic Chemistry*, 2nd ed. ed.; Wiley-Interscience: New York, 2003.
- (25) Gärnerman, D. *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*; Chapman & Hall: New York, 1997.
- (26) Gelman, A.; Roberts, G. O.; Gilks, W. R. Efficient Metropolis jumping rules. In *Bayesian Statistics 5*; Bernardo, J. M., Berger, J. O., Dawid, A. P., Smith, A. F. M., Eds.; Oxford University Press: Oxford, 1996.
- (27) Akaike, H. In *Information theory and an extension of the maximum likelihood principle*, Second International Symposium on Information Theory, Budapest, 1973; Petrov, B. N., Czaki, F., Eds.; Budapest, 1973.
- (28) *Guidance on information requirements and chemical safety assessment Chapter R.11: PBT and vPvB Assessment (Draft Version 2.0)*; European Chemicals Agency (ECHA): Helsinki, 2014.
- (29) Jager, T.; Vermeire, T. G.; Rikken, M. G. J.; van der Poel, P. Opportunities for a probabilistic risk assessment of chemicals in the European Union. *Chemosphere* **2001**, *43*, 257–264.