

This document is confidential and is proprietary to the American Chemical Society and its authors. Do not copy or disclose without written permission. If you have received this item in error, notify the sender and delete all copies.

Deriving persistence indicators from regulatory water-sediment studies – Opportunities and limitations in OECD 308 data

Journal:	<i>Environmental Science & Technology</i>
Manuscript ID:	Draft
Manuscript Type:	Policy Analysis
Date Submitted by the Author:	n/a
Complete List of Authors:	Honti, Mark; Hungarian Academy of Sciences, MTA-BME Water Research Group Fenner, Kathrin; Eawag/ETH Zürich, Environmental Chemistry

SCHOLARONE™
Manuscripts

1 Deriving persistence indicators from regulatory water-sediment studies – 2 Opportunities and limitations in OECD 308 data

3 *Mark Honti^a, Kathrin Fenner^{b,c,*}*

4 ^a MTA-BME Water Research Group, Hungarian Academy of Sciences, Budapest, Hungary

5 ^b Department of Environmental Systems Science (D-USYS), ETH Zürich, Zürich, Switzerland

6 ^c Eawag, Swiss Federal Institute of Aquatic Science and Technology, 8600 Dübendorf, Switzerland

7 * Corresponding author: kathrin.fenner@eawag.ch, Tel + 41 58 765 5085

8

9 **Abstract**

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

31 **Introduction**

32 The degradation of chemicals in different environmental compartments is one of the major
33 determinants of their environmental fate and therefore plays a crucial role in regulatory decision-
34 making. On the one hand, degradation half-lives are compared to half-life criteria for persistence
35 assessment in the context of different hazard assessment schemes ¹. On the other hand, degradation
36 rate constants are needed for higher-tier exposure modeling such as, for instance, in the FOCUS
37 models used for the risk assessment of plant protection products and veterinary pharmaceuticals ^{2,3},
38 or in the EUSES model used for risk assessment of industrial chemicals and biocides ⁴. Simulation
39 studies, such as those described in the Organisation for Economic Co-operation and Development
40 (OECD) guidelines (OECD 303 for aerobic sewage treatment (activated sludge, biofilms), OECD
41 307 for soil, OECD 308 for water-sediment systems, OECD 309 for surface waters and OECD 314
42 for wastewater treatment) are intended to produce the kind of biodegradation kinetic data that can
43 support modeling and persistence assessment at the higher tiers of regulatory risk assessment.
44 However, there are still a considerable number of open questions with respect to the performance,
45 evaluation and interpretation of those studies. Most importantly, the terms persistence and half-life
46 are not consistently defined in different regulatory frameworks and clear guidance on how they
47 should be derived from simulation study data is lacking ^{5,6}.

48 The OECD 308 guideline on “Aerobic and Anaerobic Transformation in Aquatic Sediment
49 Systems” was adopted in 2002 ⁷. It is an integral part of tiered testing strategies in different
50 legislative frameworks for the environmental risk assessment of chemicals. The experiment uses a
51 stagnant and thus fully settled water-sediment system with the parent compound initially spiked into
52 the water phase. Due to the requirement on stagnancy, the system can only be sampled by sacrificial
53 sampling, which then requires up to 30 replicates for one sediment to be ran in parallel over a time
54 period of up to 150 days. For regulatory purposes, two such studies, one with a high and one with a
55 low organic carbon content sediment, respectively, are required. The results of these highly
56 laborious and costly studies (approx. 100'000 Euros/compound) are interpreted in various ways,
57 depending on regulatory context. While their primary use in all frameworks is to derive persistence
58 indicators for hazard assessment, or degradation rate constants for exposure assessment/modeling,
59 there are also other endpoints derived from OECD 308 studies. These are the identity, amount and
60 persistence of transformation products formed, the extent of mineralization, the formation of non-
61 extractable residues, and the extent of mass transferred to the sediment. The last is used as a trigger
62 for sediment ecotoxicity testing of human pharmaceuticals ⁸.

63 Several shortcomings of the OECD 308 guideline have been identified and discussed over the years
64 in workshops ⁹⁻¹¹, Q&A documents ¹² and the scientific literature ¹³⁻¹⁶. Major points of criticism are

65 related to the lack of environmental relevance (e.g., sediment depth, stagnant conditions, extensive
66 formation of non-extractable residues etc.), and the lack of clarity on how to interpret the study
67 outcomes with respect to the main objective, i.e., to assess a chemical's degradation at the water-
68 sediment interface. The latter difficulties stem from the fact that OECD 308 assesses degradation in
69 a two-phase system where phase transfer processes and degradation happen simultaneously.

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

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

86 In contrast, $DegT_{50,system}$, which is the time until 50% of the parent chemical has disappeared from
87 the system and which is also directly observable from the data ¹⁵, is clearly a degradation indicator
88 since the parent chemical can only disappear from the total system through transformation to either
89 metabolites or non-extractable residues. Practically, this indicator is even simpler to access than the
90 previous two since it is determined from the amount of extractable parent in the total system and
91 hence does not require the separation of the sediment and water phase, which is hard to standardize.
92 Thus, $DegT_{50,system}$ is a potential candidate for an easily accessible indicator of persistence at the
93 water-sediment interface. However, it should certainly not be used as half-life for exposure
94 modeling where different environmental compartments are typically treated separately and hence
95 compartment-specific degradation rate constants are required.

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

98 that are fully independent of phase transfer processes and are potentially suitable for both
99 persistence assessment and exposure modeling²⁰. However, they are not directly observable from
100 the data and therefore have to be derived by inverse modeling. Such decomposition of the
101 measurement data necessarily introduces considerable uncertainty into the estimated kinetic
102 parameters due to model uncertainty and fundamental correlations between the estimated phase
103 transfer and degradation rate constants. Due to the complexity of simulating the OECD 308 system,
104 inverse modeling has been rarely attempted. Radke et al. (2009)²¹ carried out the estimation of
105 transformation rate constants on a variant of OECD 308, but neither discussed model uncertainty
106 nor assessed corollary parameter uncertainty.

107 Altogether, there is an obvious trade-off between conceptual robustness and uncertainty in the
108 derivation of degradation measures from OECD 308 data. The objective of this paper therefore is
109 to make this dilemma transparent and quantitative such as to deliver a sound basis for deciding
110 which indicator is best used in which regulatory context and what the “best practice” should be for
111 using this indicator. To reach this goal, we (i) develop an inverse modeling framework to quantify
112 $\text{DegT}_{50,w}$ and $\text{DegT}_{50,\text{sed}}$ values and their uncertainties for different model hypotheses on the
113 processes taking place in the water-sediment system, and apply this framework to 23 pesticides and
114 pharmaceuticals spanning a wide spectrum of degradation and sorption behavior; (ii) explore the
115 uncertainties of $\text{DegT}_{50,\text{system}}$ as a function of different experimental systems, and (iii) based on the
116 results of (i) and (ii), give recommendations on the suitability and best-practice of using
117 $\text{DegT}_{50,\text{system}}$ and $\text{DegT}_{50,w}$ and $\text{DegT}_{50,\text{sed}}$, respectively, as measures for persistence assessment and
118 exposure modeling.

119 **Methods**

120 *Data*

121 We analyzed 41 experimental OECD 308 datasets for 23 different pharmaceuticals ($n=7$) and
122 pesticides ($n=16$). Compounds were selected based on data availability and to adequately cover the
123 range of sorption behavior (i.e., organic carbon-water partition coefficients) and persistence (i.e.,
124 $\text{DegT}_{50,\text{system}}$ values from OECD 308 testing) represented in the Pesticide Properties DataBase
125 (<http://sitem.herts.ac.uk/aeru/ppdb/>). For pesticides, the OECD 308 data were read from Draft
126 Assessment Reports (DARs) made publically available by the European Food Safety Authority²².
127 For the seven pharmaceuticals, including three anonymous compounds, the data were provided by
128 two pharmaceutical companies. For the majority of compounds (18 out of 23) data from two
129 experiments were available, one carried out with a sediment possessing a higher and one with a
130 lower organic matter content as required by the 308 guideline. A list of compounds and available
131 data sets is given in Table S1 in the Supporting Information (SI).

132 *Model structure*

133 The stagnant water-sediment setup in OECD 308 experiments leads to strong heterogeneity within
134 the small experimental vessels. The lack of stirring means that diffusion governs all transport
135 processes, which may result in concentration gradients along the 2-3 cm deep sediment layer.
136 Specifically, we experimentally confirmed that, except for a very shallow surface layer, the
137 sediment is anoxic because of O₂ consumption taking place in the sediment and the limited supply
138 through the stagnant water body (see Fig. S1 in the SI). Besides this, there is a slowly developing
139 concentration gradient of the parent compound and the metabolites as they penetrate deeper into the
140 sediment via diffusion in porewater. This diffusion is slowed-down to a compound-specific degree
141 by sorption to the immobile sediment particles. As the above-mentioned gradients are packed into a
142 shallow vertical space, the geometry of the actual experimental systems may exert a considerable
143 influence on the compounds' behavior. Despite the obvious importance of these boundary
144 conditions, the OECD 308 guideline does not specify strict rules for geometrical parameters
145 (sediment depth can vary between 2-3 cm, water:sediment volume ratio between 3-4). While a
146 proper mathematical model of the OECD 308 system thus has to account for the heterogeneities in
147 the sediment to produce a good fit to measured data, no complementary data on concentration
148 gradients are typically measured in OECD 308 studies.

149 We therefore created four different but closely related mechanistic models that represent the same
150 OECD 308 experimental system with different hypotheses on compound fate in the sediment.
151 Several assumptions applied to all versions, as follows. All models used two compartments:
152 chemicals inside the system were assumed to be either in dissolved phase in the water column (i.e.,
153 neglecting sorption to dissolved organic carbon) or in sorbed or dissolved state in the sediment
154 compartment. Generated CO₂ was assumed to be trapped outside the system. All transformations
155 were assumed to follow first-order kinetics in both compartments. Diffusion into the sediment was
156 described by Fick's law. Sorption equilibrium was assumed to be reached instantaneously.
157 Transformation pathways were identical in all versions. The parent compound initially added to the
158 water compartment could be either transported to the sediment or transformed to undifferentiated
159 metabolites in the water column. In the sediment, the parent compound could be transformed either
160 to metabolites or non-extractable residues (NER), metabolites could be transformed to CO₂ or NER.
161 The four model versions differ in the chemical and physical fate of the compounds in the sediment
162 compartment (see Fig. S2 in the SI). In model A, we assumed instantaneous mixing along the entire
163 depth of the sediment for both parent compound and metabolite(s) when entering the sediment
164 column. In this case, the sediment compartment itself was treated as a fully mixed reactor, i.e.,

165 transformation processes were assumed to take place synchronously and to the same extent
166 throughout the entire sediment. This is in line with recommendations given in the FOCUS Guidance
167 Document on Estimating Persistence and Degradation Kinetics from Environmental Fate Studies²⁰
168 at level P-II. In model B, we accounted for slow diffusion into the sediment, calculating an effective
169 sediment depth that increases with time as suggested in Appendix IX of the FOCUS Guidance
170 Document²⁰. Transformation was thus limited to the actual penetration depth. As the sediment is
171 typically anaerobic except for the thin surface layer (see Fig. S1 in the SI), this implicitly assumes
172 equal aerobic and anaerobic transformation rate constants. Therefore, in versions C and D, diffusion
173 into the sediment was explicitly simulated by splitting the sediment column into 4 horizontal
174 compartments. In version C, transformation was assumed to be exclusively aerobic and to take
175 place in the uppermost boundary layer only. In version D, transformation was allowed in the deeper
176 anoxic layers too, albeit at a reduced rate. More details on model versions A to D and all model
177 equations are given in the SI.

178 *Parameter inference*

179 Although our 4 model versions are still simplified descriptions of the true system, they already
180 possess many degrees of freedom with regard to calibration. Version D has 15 model parameters
181 (including 6 transformation rate constants), 6 initial values and 1 error parameter. Versions A, B,
182 and C have 14 model parameters, 6 initial values and again 1 error parameter. These relatively large
183 parameter sets would be difficult to calibrate based on experimental data alone. The measurements
184 follow certain compound pools in the system, while parameters govern processes that move the
185 compounds between the measured pools. Due to the complexity of the models several processes can
186 compensate for each other's effect, which leads to identifiability problems. These in turn lead to
187 infinite sets of totally different model parameter values producing a similarly good fit to the data.

188 We used Bayesian parameter inference to overcome the identifiability problems in a transparent and
189 formal statistical way. Bayesian parameter inference has already been used in modeling
190 transformation kinetics^{18,23}, but exclusively with relatively simple, single-compartment models and
191 non-informative priors. Here, existing information that is independent of the measurements
192 themselves was formulated as prior probability distributions for the parameters (Tab. S3 in SI).
193 Organic carbon-water partition coefficients (K_{oc}) determined in different soils were used to define
194 priors for K_{oc} of the parent compound and for the ratio of K_{oc} of the parent and the most relevant
195 metabolite(s) (Δ). Experiment-specific data were used for the organic content of the sediment. For
196 the remaining parameters, we used expert estimations where available (e.g., diffusion coefficient
197 was set to mean=0.5 and standard deviation=0.5 cm² d⁻¹ based on²⁴) and broad general priors
198 according to our expectations where there was no prior information whatsoever.

199 Parameters were then inferred following Bayes' rule: the posterior probability of a certain
200 parameter set was proportional to the prior probability of the set times the likelihood of the set
201 given the experimental data (see, for instance, ²³). The likelihood was calculated with a simple
202 frequentist error model that assumed independent and identically distributed normal errors. The
203 standard deviation of the error was calibrated together with other model parameters ²³. While it
204 would have been possible to only look for the maximum posterior solution that best fits the data, we
205 considered the uncertainty of our parameter estimates at least as important as the best values. To
206 discover the parameter posteriors we used Markov Chain Monte Carlo (MCMC) sampling using the
207 traditional Metropolis algorithm ²⁵. The covariance matrix of the proposal distribution was tuned
208 during the burn-in phase to keep acceptance between 15 and 40% ²⁶. Three MCMC chains were
209 generated to ensure convergence and stability, each having a total length of 200,000 rounds with
210 100,000 rounds dedicated for burn-in. Samples were thinned by a factor of 5 to reduce serial
211 correlation.

212 *Analyses of parameter robustness*

213 Several sensitivity analyses were carried out to discover the robustness of various persistence
214 indicators (Tab. 1). The necessity of the Bayesian parameter inference procedure (involving
215 independent information in addition to the experimental data in the form of informative prior
216 distributions) was checked by attempting a parameter uncertainty analysis with fully vague priors
217 for each parameter (uniform distribution on the $[0, \infty[$ interval or over the physically meaningful
218 range) for model version D. The outcome was compared to the results of version D with the
219 standard priors. The sensitivity of transformation rate constants on the (informative) prior
220 distributions was tested by comparing the output of model version D (the most flexible) using the
221 standard and a stricter set of prior distributions for K_{oc} of the parent and metabolites and diffusion
222 coefficients. The importance of describing diffusion in the sediment was tested by comparing the
223 quality of fit in model version A (assumes mixed sediment) vs. versions B, C and D (assume
224 gradual diffusion into sediment). The role of hypotheses about the place of transformation in the
225 sediment was tested by comparing transformation rate constants derived from model versions B and
226 C, because these model versions formulated the most extreme hypotheses about the relative
227 importance of aerobic and anaerobic degradation. Finally, we calibrated model version D on
228 individual experiments alone and also in combination for compounds possessing two datasets from
229 different sediments. In this way, we could test whether the experimental data provided any evidence
230 to refute the universality hypothesis (i.e., that the transformation rate constant was the same in both
231 sediments), and, thus, indirectly test the validity of our assumption that transformation rate
232 constants differ across sediments for an individual compound.

233 Table 1. Summary of comparative analyses targeted at parameter robustness

Model	Priors	Sediments	Target of analysis
D	vague vs. standard vs. narrow	individual	necessity of Bayesian calibration & 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

234

235 *Analysis on the influence of system geometry on DegT_{50,system}*

236 The OECD 308 guideline formulates somewhat soft criteria regarding the geometry of the
 237 experimental systems. Sediment depth must be between 2-3 cm, the water:sediment volumetric ratio
 238 between 3-4. This can ultimately result in water column heights varying between 6 and 12 cm.
 239 These varying system geometries can influence apparent persistence in the experimental system.
 240 For instance, for compounds that are transformed faster in the sediment than in the water column,
 241 the system can theoretically be manipulated to show a relatively shorter DegT_{50,system} by using a
 242 lower water:sediment ratio. To investigate the possible magnitude of this influence we ran a series
 243 of numerical experiments under the smallest and largest allowed water:sediment ratios with model
 244 versions B, C, and D and the maximum posterior likelihood parameters of our experimental
 245 compounds. DegT_{50,system} was calculated as disappearance half-life of the parent compound using
 246 the single first order (SFO) method suggested in the FOCUS guideline²⁰.

247 **Results and discussion**248 *Necessity of Bayesian parameter inference & robustness against priors*

249 The power of Bayesian parameter inference is depending on the prior distributions. Using vague
 250 priors diminishes the advantage of incorporating existing knowledge and weakens the ability to
 251 resolve identification problems. Using very confident (narrow) priors brings the danger of relying
 252 strongly on potentially wrong information, which would introduce a significant bias in the
 253 parameters.

254 Model version D with vague priors could achieve equally good or even better fits to the
 255 experimental data than the same model coupled with the standard priors. This was expected
 256 because Bayesian parameter inference with informative priors actually forces a compromise
 257 between model fit and compliance to the expectations about the parameters. Thus, from a strictly
 258 mathematical point of view, calibrating a process-based model on OECD 308 data requires neither

259 informative priors nor a Bayesian calibration procedure. However, due to the high number of
260 interacting model parameters, it is not guaranteed that the calibrated parameter values conform to
261 the modeller's knowledge or expectations. Compared to the standard prior set, vague priors
262 introduced convergence problems into MCMC sampling and increased the uncertainty (in terms of
263 bias and variability) of estimated parameters. Inferred parameters often violated physical
264 constraints, expectations based on common sense, and recommended OECD 308 system
265 parameters. Posterior transformation rates were more robustly estimated, but sometimes reached
266 extreme values (up to 10^4 d^{-1} ; Fig. S3 in SI).

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

273 *Robustness against hypotheses on diffusion inside sediment*

274 The comparison of maximal posterior probabilities for model versions A versus B, C, and D
275 revealed that model version A failed to attain a good fit to the data in many more cases than the
276 more complicated model versions B-D (Fig. S5 in SI). The total log posterior probability was
277 significantly lower for the entire dataset for model version A than for the others (Fig. S6 in SI).
278 Model version D was the most successful due to its highest structural flexibility. This comparison
279 proved that the simulation of the gradual penetration into the sediment was a necessary feature to
280 describe the experimental data with high fidelity. This finding is actually in line with the
281 recommendations of Annex IX of the FOCUS guidance document to consider the effective depth a
282 substance has moved into the sediment when calculating the expected fraction in the sediment to
283 restrain parameter estimation²⁰. However, when reviewing the evaluations of OECD 308 data
284 presented in the pesticide DARs, we found that these recommendations are not implemented in
285 practice.

286 *Robustness against hypotheses on transformation inside sediment*

287 Version A was proven to be a true outlier in terms of quality of fit due to the lacking description of
288 diffusion inside the sediment. The other versions B, C, and D achieved roughly similar log posterior
289 probability scores despite their different assumptions on transformation mechanism. As version D
290 was capable of behaving like either versions B and C, it provided the best fit for most compounds.
291 However, the quality gain by its additional parameter (∂ , which provides version D with the ability

292 to behave like B [$\partial=1$] or C [$\partial=0$] was not enough to yield a much better, i.e., lower, Akaike
293 Information Criterion (AIC)²⁷ score (B: 11756; C: 11533; D: 11245). This indicated that the
294 experimental data did not provide any evidence about where transformation was actually taking
295 place along the sediment profile. Thus, there was not enough information to decide between the
296 mutually exclusive hypotheses of version B and C, or the compromise hypothesis manifested in
297 version D.

298 The consequences of this equifinality for the derived transformation rate constants were severe (Fig.
299 1). The assumption in C relative to B was that transformation in the sediment took place in a thin
300 surface layer, which enclosed only a small fraction of the total sediment pool of the compound once
301 it had started to diffuse into the sediment. To produce the observed amount of metabolites, CO₂, and
302 NER, sediment transformation rate constants were estimated to be up to an order of magnitude
303 higher in version C than in version B. Values for version D were in between the two extremes. In
304 contrast, since the water compartments of the four model versions were the same, transformation
305 rate constants in water were hardly affected by differences in model versions B, C and D.

306 *Universality of transformation rate constants among sediments*

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

315 Interestingly, although joint calibration could certainly not surpass the quality of fit from the
316 solutions that considered a single experiment alone, the fit was not significantly worse than the fits
317 obtained with the individual experiments in most cases (12 out of 15 compounds had less than 25%
318 difference in the calibrated standard deviation of model error). At the same time, universal
319 transformation rate constants differed significantly for both water and sediment ($p < 0.013$ in all
320 cases) from the results of individual experiments for several compounds (Fig. S7 in SI). This
321 suggests that there is no clear evidence in the data about transformation rate constants being
322 sediment-specific, because the hypothesis on the universality of the transformation rate constants is
323 performance-wise equivalent to the assumption of sediment-specific transformation rate constants.

324

325 *Robustness against variations in system geometry*

326 The allowed variation in water:sediment ratio can indeed influence $\text{DegT}_{50,\text{system}}$ (Fig. 2). There was
 327 up to 40% relative difference in $\text{DegT}_{50,\text{system}}$ between the ‘thin’ (water:sediment=4) and ‘thick’
 328 (water:sediment=3) sediment systems until a few hundred days of $\text{DegT}_{50,\text{system}}$, which is the
 329 relevant range considering the typical duration of OECD 308 experiments. Afterwards the
 330 difference grew, sometimes exceeding 100%. Interestingly, neither the sign of difference, nor the
 331 model version in which the difference showed up followed a clear pattern, but $\text{DegT}_{50,\text{system}}$ was the
 332 most extreme for the $Z_w=6\text{ cm}$, $Z_s=2\text{ cm}$ and $Z_w=12\text{ cm}$, $Z_s=3\text{ cm}$ pairs in all cases. This suggests
 333 that without knowing anything about actual system geometry, $\text{DegT}_{50,\text{system}}$ may, on average, have at
 334 least 40% a priori uncertainty due to the sloppily defined experimental geometry standards. This
 335 can influence the classification of compounds as (non)persistent: in our dataset of 41 experiments
 336 and 3 model versions (B,C,D), $\text{DegT}_{50,\text{system}}$ crossed the threshold value for persistence in water of
 337 40 days for 7 cases, and the threshold values for persistence in sediment of 120 days for 2 cases,
 338 within the allowed variation of system geometry (Fig. 2).

339 **Consequences for using OECD 308 data in a regulatory context**

340 There is a disparity between the uncertainty and the environmental relevance of persistence
 341 indicators derivable from OECD 308 data (Tab. 2). Indicators such as $\text{DegT}_{50,\text{system}}$, $\text{DT}_{50,\text{w}}$ and
 342 $\text{DT}_{50,\text{sed}}$ are directly extractable from the data with limited uncertainty, but they are specific to the
 343 experimental system. Less system-specific indicators (such as $\text{DegT}_{50,\text{w}}$ and $\text{DegT}_{50,\text{sed}}$) require a
 344 rather involved inverse modeling approach to extract and suffer from considerable uncertainty.

345 Table 2. Comparative analysis of different persistence indicators derivable from OECD 308 data
 346 with respect to conceptual robustness and uncertainty

	Indicators		
	$\text{DegT}_{50,\text{system}}$	$\text{DegT}_{50,\text{w}}$	$\text{DegT}_{50,\text{sed}}$
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.	

347

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

353 As pointed out in the introduction, phase-specific apparent persistence indicators ($\text{DT}_{50,\text{w}}$ and
354 $\text{DT}_{50,\text{sed}}$) cannot be considered as robust degradation measures because they lump together
355 transformation and phase transfer. This is confirmed by the comparison of $\text{DegT}_{50,\text{w}}$ values
356 estimated in this study with $\text{DT}_{50,\text{w}}$ values derived from the same data as presented in Fig. 3.
357 $\text{DegT}_{50,\text{w}}$ was found to be always much longer than $\text{DT}_{50,\text{w}}$. This indicated that $\text{DT}_{50,\text{w}}$ was actually
358 dominated by diffusion, and transformation only played a secondary role. At the same time,
359 $\text{DegT}_{50,\text{w}}$ was always smaller than or equal to chemical hydrolysis half-lives reported in the EFSA
360 DAR documents²², which confirmed that transformation in the water column of OECD 308 studies
361 also reflects biological processes (Fig. 3). It is of further interest to note that using $\text{DegT}_{50,\text{w}}$ instead
362 of $\text{DT}_{50,\text{w}}$ for comparison against the persistence threshold for freshwater (40 d) as suggested in the
363 REACH draft guidance for PBT assessment²⁸, results in a promotion from non-persistent to
364 persistent for the majority of substances investigated (Fig. 3).

365 Less system-specific indicators that exclude phase transfer ($\text{DegT}_{50,\text{w}}$ and $\text{DegT}_{50,\text{sed}}$) are
366 conceptually more robust, but this advantage comes at the cost of elevated uncertainty. According
367 to our results, $\text{DegT}_{50,\text{w}}$ was identifiable and similar for all models, most probably due to the
368 identical description of the water phase in all model versions. Even so, however, the estimated
369 $\text{DegT}_{50,\text{w}}$ values were still subject to large parametric uncertainty (CVs in the range of 0.5-3.0)
370 resulting from inverse modeling. On top of this parametric uncertainty, $\text{DegT}_{50,\text{sed}}$ showed a strong
371 dependence on the assumptions on transformation mechanisms in the sediment. Available data did
372 not provide a way to decide between the competing assumptions on the compounds' fate in the
373 sediment, so the different (uncertain) estimates would need to be merged in an even more uncertain
374 composite estimate to get a real picture about the uncertainty of $\text{DegT}_{50,\text{sed}}$. Moreover, the
375 universality of $\text{DegT}_{50,\text{sed}}$ values among different sediments could neither be proven nor refuted.
376 The competing hypotheses of $\text{DegT}_{50,\text{sed}}$ values being sediment-specific or universal for a certain
377 compound delivered statistically equivalent fits to the data. However, the two hypotheses resulted in
378 significantly different estimates, which pointed towards a so far not quantified uncertainty.

379 Despite our sophisticated inverse modeling approach, the data available from experiments carried
380 out according to the OECD 308 guideline were thus insufficient to derive persistence indicators that

381 had both acceptable levels of robustness and uncertainty at the same time. Based on these findings,
382 we would like to outline two strategies: (i) a strategy to extract the maximal, yet defensible amount
383 of persistence information from existing OECD 308 data, and (ii) a strategy to improve our ability
384 to assess persistence at the sediment-water interface in future testing.

385 With regard to using existing OECD 308 data, we suggest that $\text{DegT}_{50,\text{system}}$ values can be used as
386 indicators of persistence at the water-sediment interface in the context of PBT assessment and for
387 the purpose of benchmarking compounds against each other. It should, however, be noted that this
388 persistence estimate is most representative of degradation in the uppermost, at least partially aerobic
389 sediment layer in a small, shallow, stagnant or slowly flowing water body. To improve the
390 conceptual robustness of using $\text{DegT}_{50,\text{system}}$ values as persistence indicators, we recommend to
391 either revise the existing OECD 308 guideline to further restrict the allowed variability in system
392 geometry or to strictly enforce reporting of the applied system geometry (i.e., the test vessel's inner
393 diameter, the heights of the water and sediment column, and sediment dry weight used). Finally,
394 proper usage of $\text{DegT}_{50,\text{system}}$ values derived from OECD 308 data as persistence indicator for PBT
395 assessment would require the definition of a persistence threshold against which these values should
396 be benchmarked.

397 For environmental exposure assessment purposes, neither $\text{DT}_{50,\text{w}}$ nor $\text{DT}_{50,\text{sed}}$ should be used, but
398 rather inverse modeling as presented here should be used to derive $\text{DegT}_{50,\text{w}}$ and $\text{DegT}_{50,\text{sed}}$ values.
399 These can then defensibly be used as input parameters into models for environmental fate and
400 exposure assessment, but their uncertainty needs to be considered. While we are confident that the
401 uncertainty of $\text{DegT}_{50,\text{w}}$ obtained from inverse modeling adequately presents the true uncertainty of
402 this parameter, the uncertainty estimates obtained for $\text{DegT}_{50,\text{sed}}$ most likely underestimate its true
403 uncertainty due to the underlying model uncertainties. Even when accounting for these
404 uncertainties, however, the Bayesian approach still leads to a clear gain in information when
405 compared to the state-of-the-art presented in the FOCUS Guidance Document on Estimating
406 Persistence and Degradation Kinetics from Environmental Fate Studies²⁰. There it is suggested to
407 use default worst-case half-lives of 1000 d for exposure modeling if fitting of the 2-compartment
408 model fails. Our results indicate that for > 90% of the substance/sediment combinations evaluated,
409 resulting half-lives in both water and sediment were lower than 1000 d.

410 It is a common critique against the OECD 308 experimental setup that it has limited direct
411 relevance for water bodies having higher water:sediment ratios, such as most real streams and lakes.
412 Nevertheless, transformation rates derived from 308 experiments could be theoretically used to
413 estimate $\text{DegT}_{50,\text{system}}$ for these different exposure situations. Our data show that the estimated

414 ‘environmental’ $\text{DegT}_{50,\text{system}}$ varies largely under the realistic range of environmental
415 water:sediment ratios and posterior parameter uncertainty (Fig. S8 in SI). For almost all
416 compounds, there are specific water:sediment ratios that either make the compound lie above or
417 below typical persistence threshold values.

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

430 **Acknowledgments**

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

437 **Supporting Information**

438 The SI contains data on the compounds and sediments used in this study, a more detailed
439 description of the model versions and various materials on the outcome of parameter inference
440 experiments. This information is available free of charge via the Internet at <http://pubs.acs.org/>.

441

442 **References**

- 443 1. Boethling, R.; Fenner, K.; Howard, P.; Klecka, G.; Madsen, T.; Snape, J. R.; Whelan, M. J.
444 Environmental Persistence of Organic Pollutants: Guidance for Development and Review of POP
445 Risk Profiles. *Int. Environ. Assess. Manag.* **2009**, *5*, 539-556.
- 446 2. *Surface Water Models and EU Registration of Plant Protection Products*; Report of the
447 work of the Regulatory Modelling Working Group on Surface Water Models of FOCUS (FORum
448 for the Coordination of pesticide fate models and their USE): 1997.
- 449 3. *Revised Guideline on Environmental Impact Assessment for Veterinary Medicinal Products*
450 *in Support of the VICH Guidelines GL6 and GL38. EMEA/CVMP/ERA/418282/2005-Rev.1*;
451 Committee for Medicinal Products for Veterinary Use (CVMP); European Medicines Agency
452 (EMA): London, 2009.
- 453 4. *Guidance on information requirements and chemical safety assessment, Chapter R.16:*
454 *Environmental Exposure Estimation*; European Chemicals Agency (ECHA): Helsinki, 2008.
- 455 5. Rauert, C.; Friesen, A.; Hermann, G.; Johncke, U.; Kehrer, A.; Neumann, M.; Prutz, I.;
456 Schonfeld, J.; Wiemann, A.; Willhaus, K.; Woltjen, J.; Duquesne, S. Proposal for a harmonised
457 PBT identification across different regulatory frameworks. *Environ. Sci. Eur.* **2014**, *26*, 9.
- 458 6. Solomon, K.; Matthies, M.; Vighi, M. Assessment of PBTs in the European Union: a critical
459 assessment of the proposed evaluation scheme with reference to plant protection products. *Environ.*
460 *Sci. Eur.* **2013**, *25*, 10.
- 461 7. *OECD guidelines for testing of chemicals; 308, Aerobic and anaerobic transformation in*
462 *aquatic sediment systems*; Organisation for Economic Co-operation and Development (OECD):
463 Paris, 2002.
- 464 8. *Guideline on the Environmental Risk Assessment of Medicinal Products for Human Use*;
465 Committee for Medicinal Products for Human Use (CHMP); European Medicines Agency
466 (EMA): London, 2006.
- 467 9. Bowmer, T.; Leopold, A. *Strategies for selecting biodegradation simulation tests and their*
468 *interpretation in persistence evaluation and risk assessment; Simulation Testing of Environmental*
469 *Persistence (STEP): Report of a two-day workshop held in Rotterdam, 4 and 5 October, 2004*;
470 Rotterdam, 2004.
- 471 10. *Workshop Report No. 17: Significance of bound residues in environmental risk assessment,*
472 *14-15 October 2009, Brussels*; European Centre for Ecotoxicology and Toxicology of Chemicals
473 (ECETOC): Brussels, 2010.

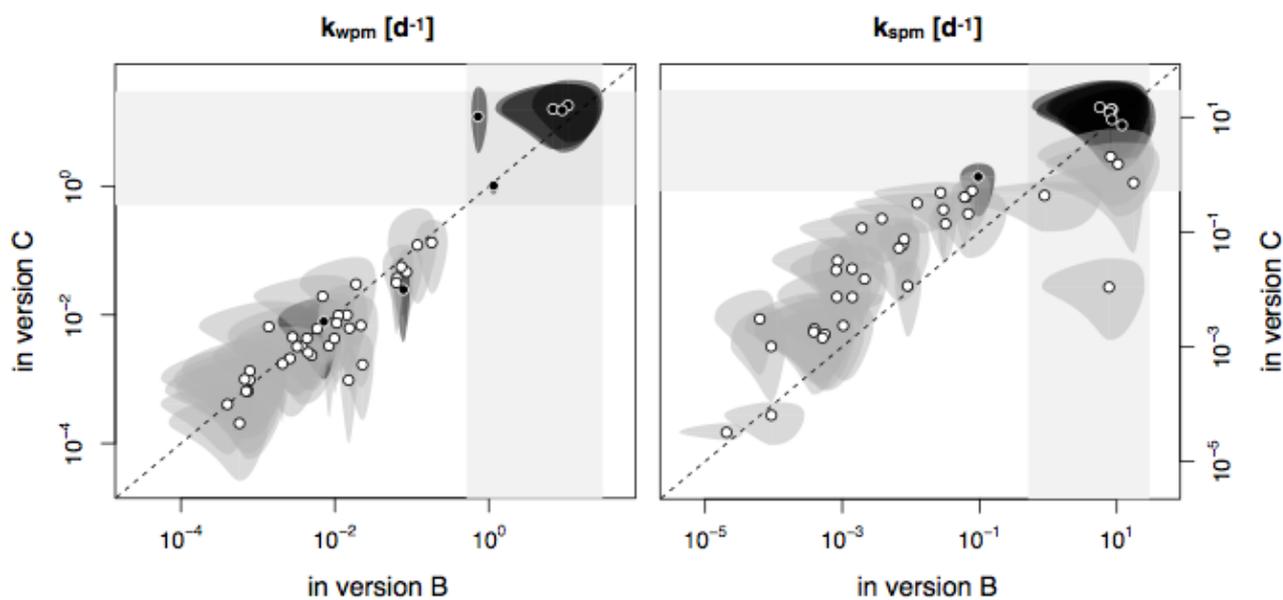
- 474 11. *Special Report No. 18: Information to be considered in a weight-of-evidence-based*
475 *PBT/vPvB assessment of chemicals*; European Centre for Ecotoxicology and Toxicology of
476 Chemicals (ECETOC): Brussels, 2014.
- 477 12. *Questions and answers on 'Guideline on the environmental risk assessment of medicinal*
478 *products for human use'*; European Medicines Agency, Committee for Medicinal Products for
479 Human Use (CHMP): London, 2011.
- 480 13. Davis, J. W.; Gonsior, S.; Marty, G.; Ariano, J. The transformation of hexabromocyclo-
481 dodecane in aerobic and anaerobic soils and aquatic sediments. *Wat. Res.* **2005**, *39*, 1075-1084.
- 482 14. Ericson, J. F. An evaluation of the OECD 308 water/sediment systems for investigating the
483 biodegradation of pharmaceuticals. *Environ. Sci. Technol.* **2007**, *41*, 5803-5811.
- 484 15. Ericson, J. F.; Smith, R. M.; Roberts, G.; Hannah, B.; Hoeger, B.; Ryan, J. Experiences with
485 the OECD 308 transformation test: A human pharmaceutical perspective. *Int. Environ. Assess.*
486 *Manag.* **2014**, *10*, 114 - 124.
- 487 16. Radke, M.; Maier, M. P. Lessons learned from water/sediment-testing of pharmaceuticals.
488 *Wat. Res.* **2014**, *55*, 63-73.
- 489 17. Kästner, M.; Nowak, K. M.; Miltner, A.; Trapp, S.; Schäffer, A. Classification and
490 modelling of non-extractable residue (NER) formation of xenobiotics in soil—a synthesis. *Crit.*
491 *Rev. Env. Sci. Tec.* **2014**, *44*, 2107-2171.
- 492 18. Loos, M.; Krauss, M.; Fenner, K., Non-extractable residue formation: Insights from kinetic
493 meta-analysis of regulatory soil simulation studies. *Environ. Sci. Technol.* **2012**, *46*, 9830-9837.
- 494 19. Matthies, M.; Witt, J.; Klasmeier, J. Determination of soil biodegradation half-lives from
495 simulation testing under aerobic laboratory conditions: A kinetic model approach. *Environ. Poll.*
496 **2008**, *156*, 99-105.
- 497 20. *Guidance Document in Estimating Persistence and Degradation Kinetics from*
498 *Environmental Fate Studies in Pesticides in EU Registration*; The Final Report of the Work Group
499 on Degradation Kinetics of FOCUS (FORum for the Coordination of pesticide fate models and their
500 USE), EC Document Reference Sanco/10058/2005, version 2.0: 2006.
- 501 21. Radke, M.; Lauwigi, C.; Heinkele, G.; Murdter, T. E.; Letzel, M. Fate of the antibiotic
502 sulfamethoxazole and its two major human metabolites in a water sediment test. *Environ. Sci.*
503 *Technol.* **2009**, *43*, 3135-3141.
- 504 22. *Rapporteur member state assessment reports submitted for the EU peer review of*
505 *activesubstances used in plant protection products*; European Food Safety Authority (EFSA), 2014.
506 <http://dar.efsa.europa.eu/dar-web/provision>.
- 507 23. Gorlitz, L.; Gao, Z. L.; Schmitt, W. Statistical analysis of chemical transformation kinetics
508 using Markov-Chain Monte Carlo methods. *Environ. Sci. Technol.* **2011**, *45*, 4429-4437.

- 509 24. Schwarzenbach, R. P.; Imboden, D. M.; Gschwend, P. M. *Environmental Organic*
510 *Chemistry*. 2nd edition ed.; Wiley-Interscience: New York, 2003.
- 511 25. Gamerman, D., *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*.
512 Chapman & Hall: New York, 1997.
- 513 26. Gelman, A.; Roberts, G. O.; Gilks, W. R. Efficient Metropolis jumping rules. In *Bayesian*
514 *Statistics 5*, Bernardo, J. M.; Berger, J. O.; Dawid, A. P.; Smith, A. F. M., Eds. Oxford University
515 Press: Oxford, 1996.
- 516 27. Akaike, H. In *Information theory and an extension of the maximum likelihood principle*,
517 Second International Symposium on Information Theory, Budapest, 1973; Petrov, B. N.; Czaki, F.,
518 Eds. Budapest, 1973.
- 519 28. *Guidance on information requirements and chemical safety assessment Chapter R.11: PBT*
520 *and vPvB Assessment (Draft Version 2.0)*; European Chemicals Agency (ECHA): Helsinki, 2014.

521

522

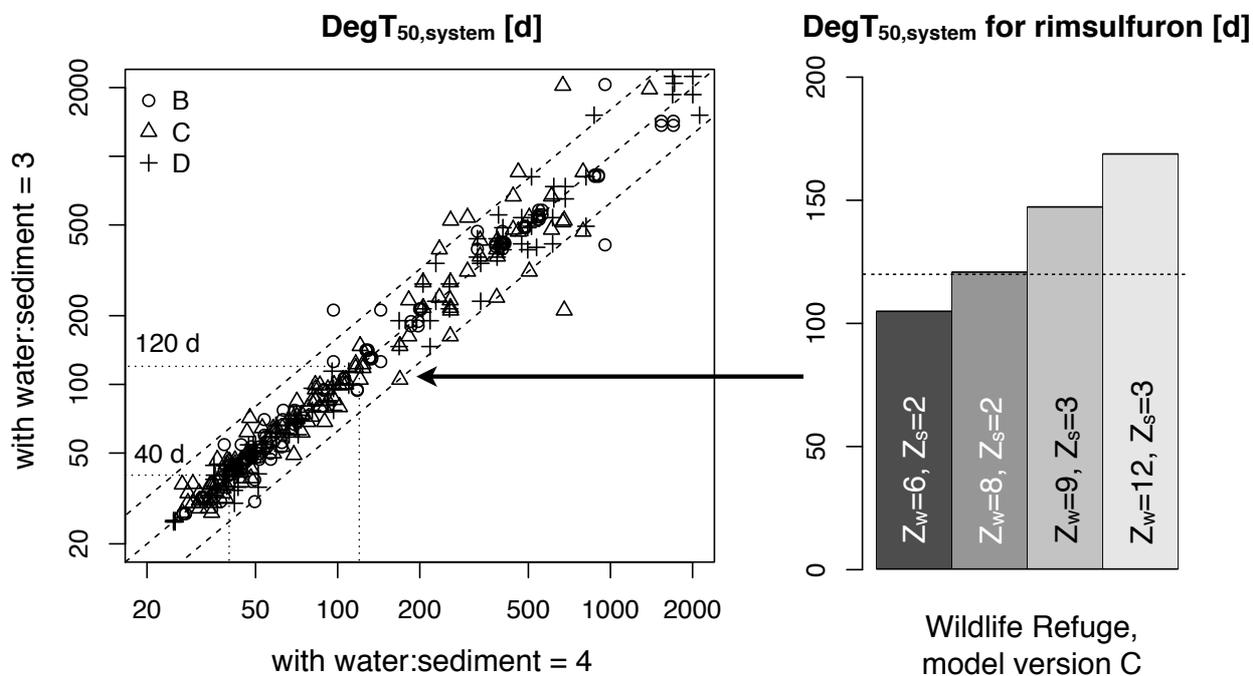
523 Fig. 1. Transformation rate constants of the parent compound in water (k_{wpm}) and in sediment (k_{spm})
524 in model version B versus version C. The grey bands indicate the 90% confidence interval of the
525 prior distribution. Black points indicate compounds that are believed to hydrolyze quickly.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



527

528

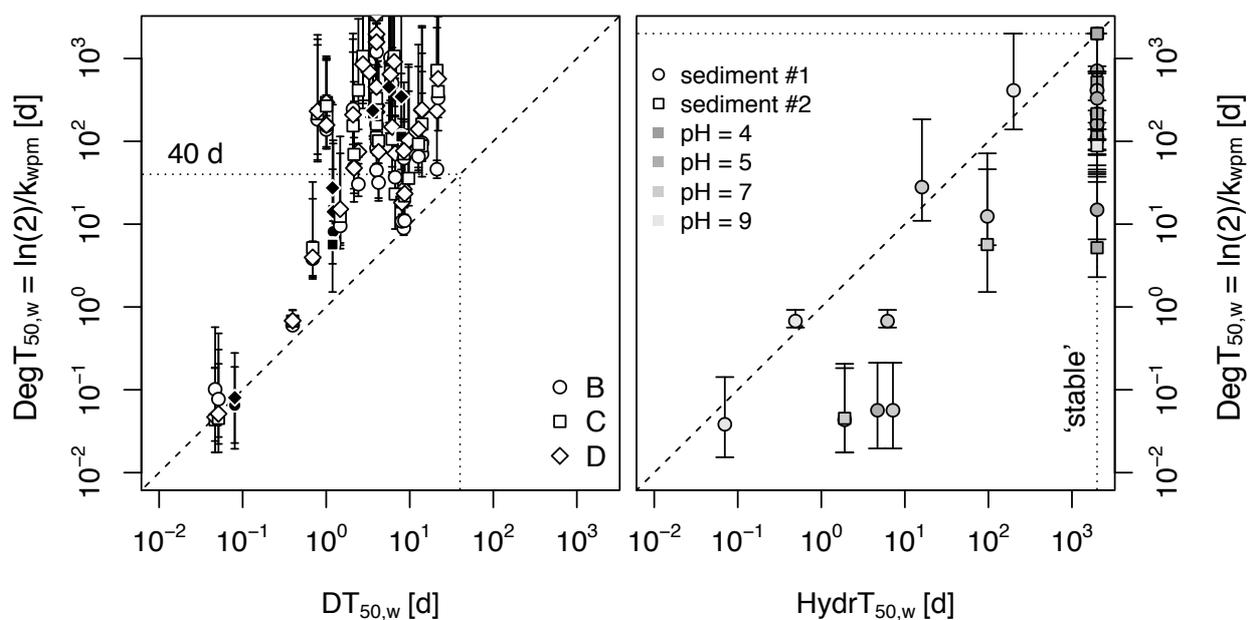
529 Fig. 2. Influence of experimental system geometry on $\text{DegT}_{50,\text{system}}$ with different model versions
 530 (left). Dashed lines show the standard deviation of all relative differences (60%). The experiment
 531 with rimsulfuron @ Wildlife Refuge with model version C is shown to illustrate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



534

535

536 Fig. 3. Comparison of different half-lives of the parent compound in water. Left: Estimated
537 transformation half-lives in water ($\text{DegT}_{50,w}$) versus $\text{DT}_{50,w}$ with model versions B, C, and D. Right:
538 Estimated transformation half-lives in water with model version D versus chemical hydrolysis half-
539 lives in water ($\text{HydrT}_{50,w}$) reported in EFSA DAR documents. Whiskers show 95% posterior
540 uncertainty intervals in both panels. Compounds found to be stable with respect to chemical
541 hydrolysis are shown at $\text{HydrT}_{50,w}=2000$ d.



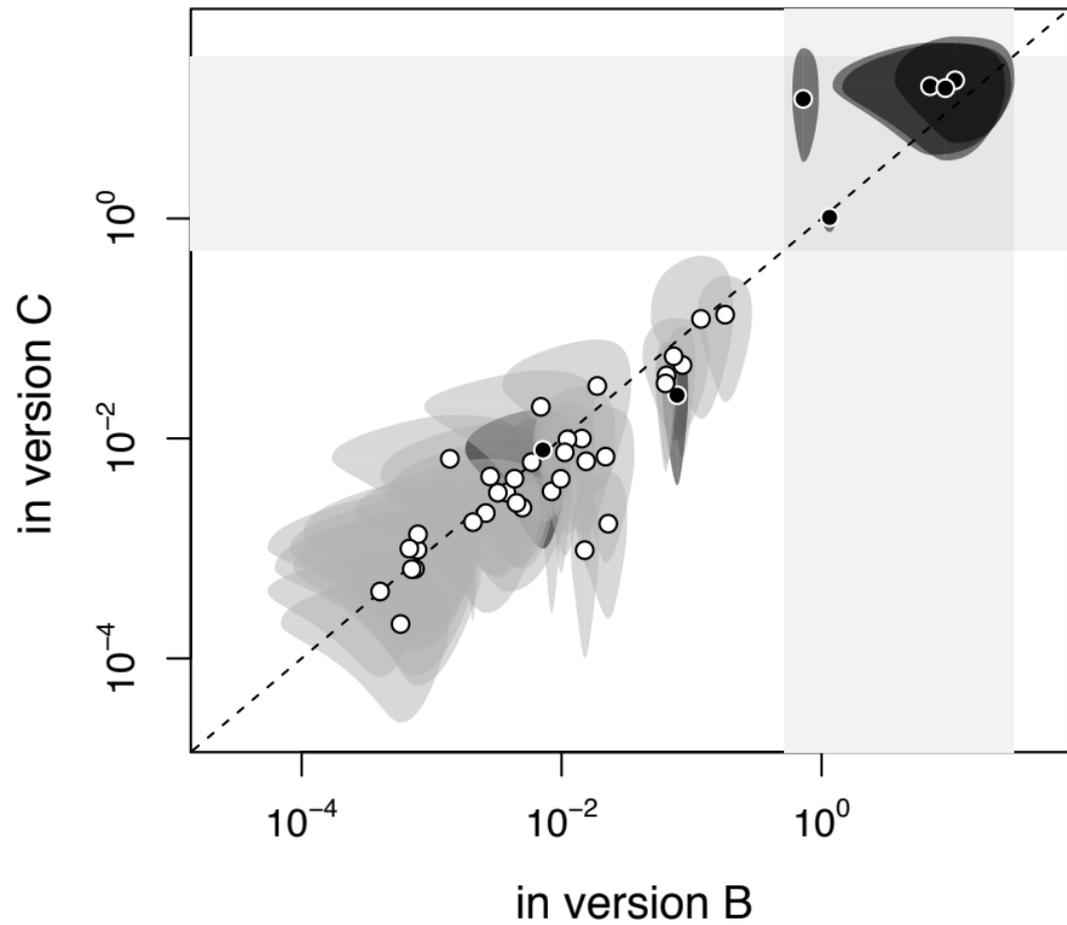
542

543

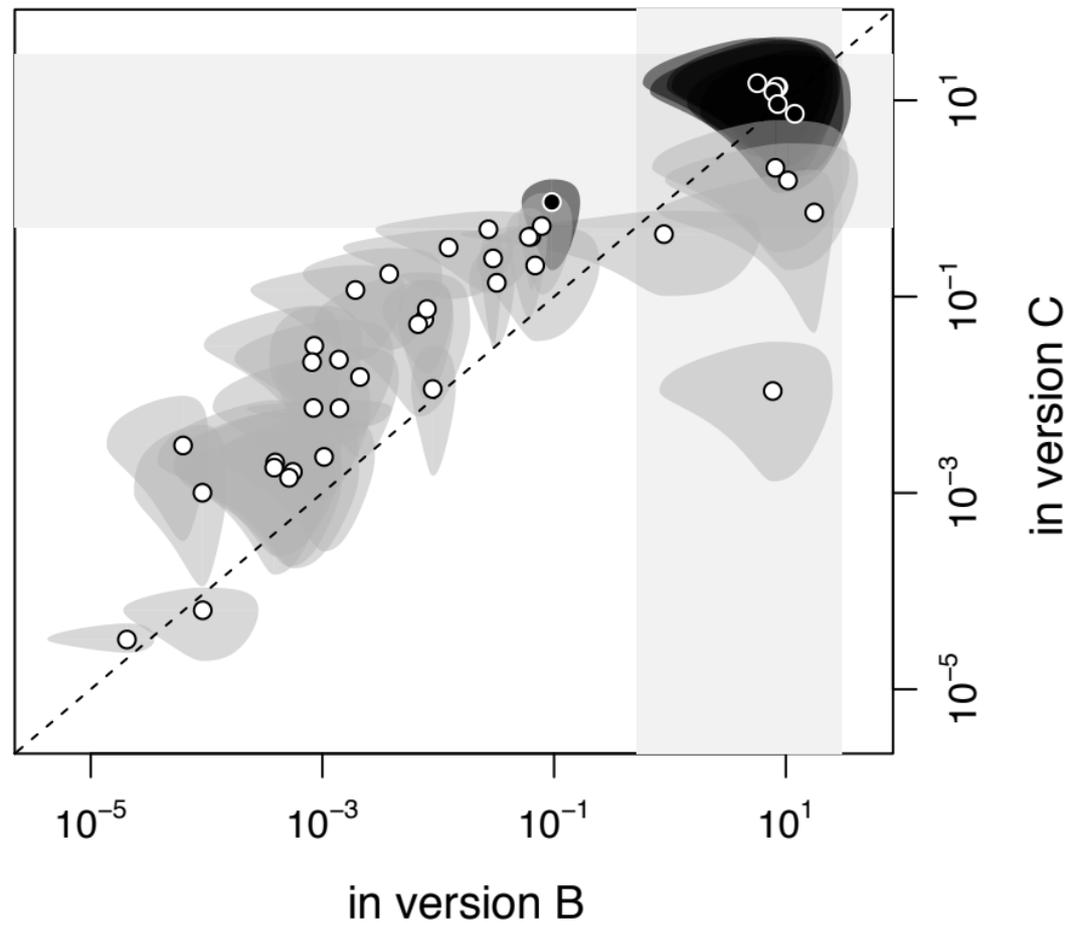
544

545

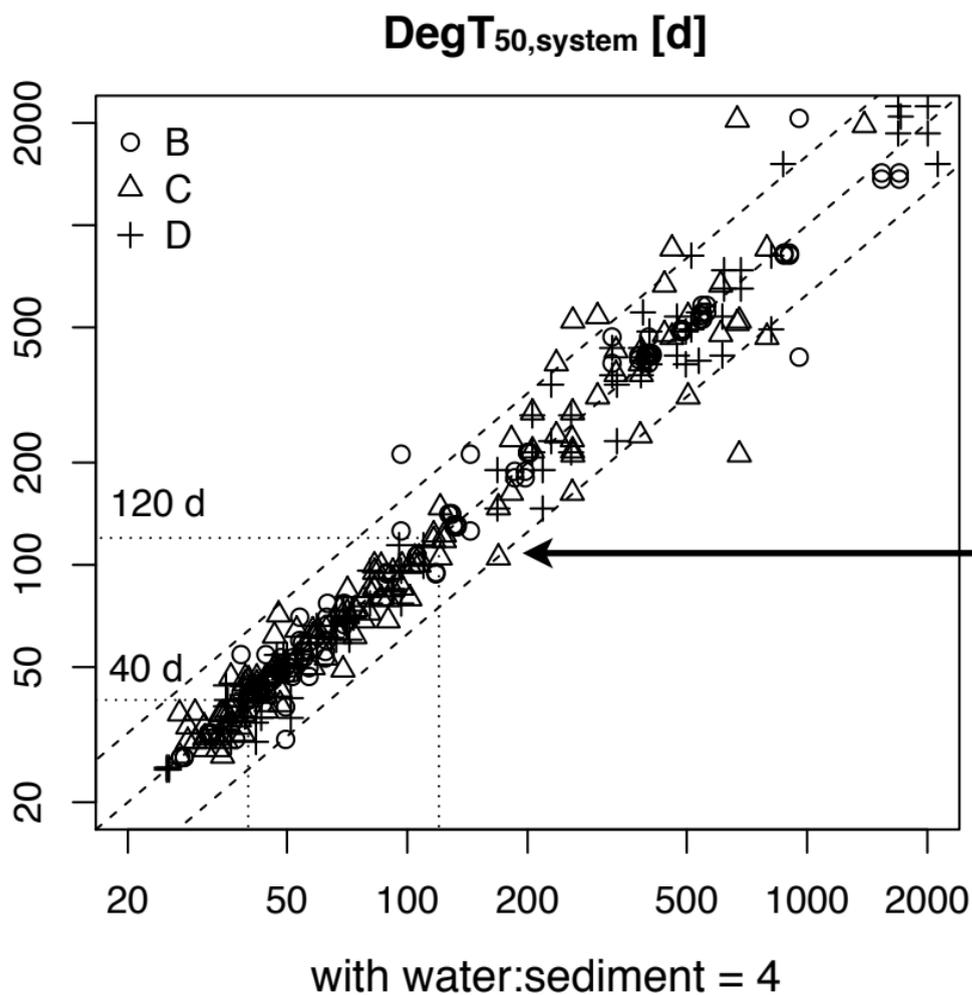
546

k_{wpm} [d^{-1}]

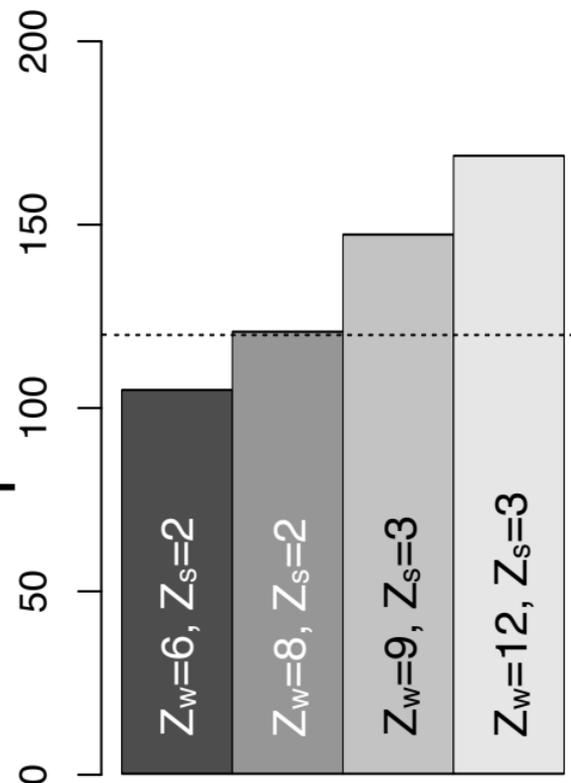
Environmental Science & Technology



with water:sediment = 3



DegT_{50,system} for rim sulfuron [d]



Wildlife Refuge,
model version C

$\text{DegT}_{50,w} = \ln(2)/k_{\text{wppm}} [\text{d}]$

