Environmental Research of the
Federal Ministry for the
Environment, Nature Conservation,
Building and Nuclear Safety

Project number: FKZ 3715 65 415 3

# Suitability of laboratory simulation tests for the identification of persistence in surface waters

by

Kathrin Fenner [1], Mark Honti [2], Christian Stamm [1], Laura Varga [2], Fabian Bischoff [1]

[1] Eawag, Department Environmental Chemistry, Überlandstrasse 133, 8600 Dübendorf, Switzerland

[2] Budapest University of Technology and Economics Department of Hydraulic and Water Resources Engineering, Műegyetem rakpart 3, 1111 Budapest, Hungary

On behalf of the German Environment Agency

November 2016

**Abstract**

Simulation tests play a crucial role in evaluating chemical biotransformation for regulatory purposes. For chemical biotransformation in surface water systems, two OECD testing guidelines are relevant: OECD 308 and OECD 309, which assess transformation at the water-sediment interface, and in the water body, respectively. Major issues with these guidelines concern the relevance of the test conditions to properly reflect degradation in actual surface water bodies, and a lack of guidance and practical tools for the derivation of actual degradation half-lives. This research project addressed both concerns. In WP I, inverse modeling approaches were applied to derive total system and compartment-specific degradation half-lives ($DegT_{50,ts}$, $DegT_{50,w}$ and $DegT_{50,sed}$) and their respective uncertainties. The persistence indicators derived from OECD 308 and 309 data all displayed significant uncertainties, i.e., typically around a factor of 2 for $DegT_{50,ts}$ and 1-2 orders of magnitude for $DegT_{50,w}$ and $DegT_{50,sed}$. These results are in conflict with the presently used rigid persistence criteria. $DegT_{50,w}$ values were found to always be higher than $DegT_{50,sed}$ values for the same compound, and accordingly OECD 309 results much more often led to a substance being classified as persistent than OECD 308 results did. The joint calibration of different experimental types via the bioavailability- and biomass-normalized $k'_{bio}$ reduced the uncertainty of $DegT_{50,w}$ and $DegT_{50,sed}$, but this reduction of uncertainty was small due to the limited extent of biotransformation observed in OECD 309 systems. It was suggested that a modified version of OECD 309 with more suspended sediment would improve the accuracy of estimating $DegT_{50,w}$ and $DegT_{50,sed}$. WP II addressed the representativeness of the laboratory-based OECD 308 and 309 simulation tests to reflect and predict the chemicals' fate in actual surface water bodies. Together with a literature review discussing the major factors influencing chemical degradation in surface water bodies and how these are reflected in different laboratory-based test systems, a case study on substance degradation in the river Rhine was conducted. In this case study monitoring data from the river Rhine were compared with experimental data from OECD 308 studies implemented in a chemical fate model to evaluate the appropriateness of OECD 308 data to reflect degradation in a large stream. It was found that application of compartment-specific half-lives did not contradict observed concentrations whereas application of $DegT_{50,ts}$ clearly overestimated degradation. Overall, based on the results of the project, the execution of two simulation studies to assess biotransformation in water-sediment systems is recommended. These should be an OECD 308 study and a 309 study with as much suspended sediment as allowed. Doing so allows deducing compartment-specific half-life indicators with reduced uncertainty, and allows considering the actual system dimensions in the field during exposure modelling as demonstrated in the Rhine case study.

**Kurzbeschreibung**

Simulationsstudien spielen bei der Bewertung der Biotransformation von Chemikalien für regulatorische Zwecke eine wichtige Rolle. Für die Biotransformation in Oberflächengewässern sind zwei OECD-Prüfrichtlinien relevant: OECD 308 und OECD 309, welche Transformation an der Wasser-Sediment-Grenzfläche beziehungsweise im Wasserkörper bewerten. Hauptkritikpunkte an diesen Richtlinien beziehen sich auf die Repräsentativität der Testbedingungen für den Abbau in realen Oberflächengewässern, sowie auf den Mangel an fehlenden Anleitungen und Instrumenten, um Abbau-Halbwertszeiten aus den Testresultaten abzuleiten. Dieses Forschungsprojekt adressiert beide Punkte. In Arbeitspaket I wurden inverse Modellierungsansätze angewendet, um Abbau-Halbwertszeiten für das Gesamtsystem und die einzelnen Kompartimente ($DegT_{50,ts}$, $DegT_{50,w}$ und $DegT_{50,sed}$) sowie ihre jeweiligen Unsicherheiten abzuleiten. Alle aus OECD 308 und 309-Daten abgeleiteten Persistenzindikatoren wiesen erhebliche Unsicherheiten auf. Diese liegen in der Regel bei einem Faktor von zwei für $DegT_{50,ts}$ und ein bis zwei Größenordnungen für $DegT_{50,w}$ und $DegT_{50,sed}$. Dieser Befund steht in Widerspruch mit der derzeitigen Verwendung von starren Persistenzkriterien. $DegT_{50,w}$-Werte lagen immer höher als $DegT_{50,sed}$-Werte für die gleiche Verbindung. Entsprechend führten auch OECD 309-Ergebnisse wesentlich häufiger zu einer Persistenz-Klassifizierung für die gleiche Substanz wie

OECD 308-Ergebnisse. Die gemeinsame Kalibrierung von verschiedenen Testsystemen mittels der Bioverfügbarkeits- und Biomasse-normalisierten $k'_{bio}$ reduzierte die Unsicherheit der $DegT_{50,w}$ und $DegT_{50,sed}$. Die Reduktion der Unsicherheit war jedoch aufgrund des begrenzten Ausmaßes der beobachteten Biotransformation in OECD 309-Systemen gering. Es wurde vorgeschlagen, dass eine modifizierte Version des OECD 309 mit mehr Schwebstoffen die Genauigkeit der Schätzung von $DegT_{50,w}$ und $DegT_{50,sed}$ verbessern würde. Arbeitspaket II adressierte die Repräsentativität der Labor-basierten OECD 308 und 309-Simulationstests, um das Schicksal von Chemikalien in tatsächlichen Oberflächengewässern zu prognostizieren. Neben einer Literaturrecherche zu den wichtigsten Einflussfaktoren auf die Biotransformation von Chemikalien in Oberflächengewässern wurde eine Fallstudie über Substanzabbau im Rhein durchgeführt. Dazu wurden Daten aus der Rheinüberwachung mit Modellvorhersagen zum Verbleib von chemischen Substanzen im Rhein verglichen. Basierend darauf wurde die Angemessenheit von Halbwertszeiten, die aus OECD 308-Daten hergeleitet worden sind, zur Beschreibung des Verbleibs von Chemikalien in einem großen Fluss wie dem Rhein bewertet. Diese Fallstudie ergab, dass die Anwendung von Kompartiments-spezifischen Halbwertzeiten nicht im Widerspruch zu beobachteten Konzentrationen steht. Die Anwendung von $DegT_{50,ts}$ jedoch überschätzte den Abbau deutlich. Insgesamt wird basierend auf den Ergebnissen des Projekts die Durchführung von zwei Simulationsstudien empfohlen, um Biotransformation in Wasser-Sediment-Systemen zu beurteilen. Dabei sollte es sich um eine OECD 308- und eine 309-Studie mit der maximal erlaubten Menge an Schwebstoffen handeln. Dadurch würde es möglich werden, Kompartiments-spezifische Halbwertszeit-Indikatoren mit reduzierter Unsicherheit abzuleiten, sowie die tatsächlichen Dimensionen des realen Oberflächengewässers in der Expositionsmodellierungen zu berücksichtigen.

# Table of contents

# List of figures

# List of tables

# List of abbreviations

| | |
|---|---|
| **API** | Active Pharmaceutical Ingredient |
| **$CO_2$** | $CO_2$ caught in traps [% of applied radioactivity] |
| **$DegT_{50,sed}$** | Degradation half-life in sediment |
| **$DegT_{50,ts}$** | Total system degradation half-life |
| **$DegT_{50,w}$** | Degradation half-life in water |
| **$DissT_{50,sed}$** | Disappearance half-life in sediment |
| **$DissT_{50,ts}$** | Total system disappearance half-life |
| **$DissT_{50,w}$** | Disappearance half-life in water |
| **$dk_{aer}$** | Relative anaerobic coefficient |
| **$dK_d$** | Relative sediment-water partitioning coefficient of transformation products |
| **DOC** | Dissolved Organic Carbon |
| **$DOC_w$** | Dissolved Organic Carbon of water column |
| **$f_{oc,sed}$** | Sediment organic carbon content |
| **$f_{om,sed}$** | Sediment organic matter content |
| **GIS** | Geographical Information System |
| **$k'_{bio}$** | Second-order biotransformation rate constant, |
| **$K_d$** | Sediment-water partitioning coefficient. |
| **$k_{hydr}$** | Rate of hydrolysis |
| **$k_{mn}$** | NER formation rates from transformation products |
| **$k_{pn}$** | NER formation rates from parent compound |
| **Koc** | Soil organic carbon-water partitioning coefficient |
| **$k_{spm}$** | First-order biotransformation rate in the sediment phase |
| **$k_{wpm}$** | First-order biotransformation rate in the water phase |
| **$\hat{M}_s$** | Solid-bound transformation products in a reactor layer |
| **$M_s$** | Transformation products in the sediment compartment |
| **$\hat{M}_w$** | Liquid-phase transformation products in a reactor layer |
| **$M_w$** | Transformation products in the water compartment |
| **NER** | Non-Extractable Residues |
| **OECD** | Organisation for Economic Cooperation and Development |

| | |
|---|---|
| **P** | Parent compound |
| **$\hat{P}_{total}$** | Total amount of the parent compound in the system |
| **PE** | Pesticide |
| **$\hat{P}_s$** | Particle-bound parent compound in a reactor layer |
| **$P_s$** | Parent compound in the sediment compartment |
| **$\hat{P}_w$** | Liquid-phase parent compound in a reactor layer |
| **$P_w$** | Parent compound in the water compartment |
| **REACH** | Registration, Evaluation, Authorisation and Restriction of Chemicals |
| **RMSE** | Root-Mean-Square-Error |
| **SETAC** | The Society of Environmental Toxicology and Chemistry |
| **SFO** | Single First-Order |
| **SOM** | Sediment Organic Matter |
| **$TOC_w$** | Total Organic Carbon concentration in water column |
| **TP** | Transformation Product |
| **TSS** | Total Suspended Solids |
| **UBA** | Umweltbundesamt |
| **WP** | Work Package |
| **WWTP** | Wastewater Treatment Plant |
| **$Z_{sed}$** | Wet sediment depth |
| **$Z_{wc}$** | Water column height |
| **$\theta$** | Porosity of wet sediment |

# Zusammenfassung

Für viele Klassen von Chemikalien ist ein bestimmter Eintrag in Oberflächengewässer im Verlaufe ihres Lebenszyklus unvermeidbar. Die diesbezüglich wichtigsten Klassen umfassen Pflanzenschutzmittel (Pestizide), welche durch Versickerung oder Abschwemmungen in Oberflächengewässer eindringen, sowie Humanarzneimittel, Biozide und Industriechemikalien, welche nur teilweise in der Abwasserbehandlung entfernt werden und somit in Gewässer eingeleitet werden. Des Weiteren können veterinärmedizinische Arzneimittel, welche durch die Aufbringung von tierischem Dünger auf den Boden gelangen von dort ebenfalls weiter in Oberflächengewässer transportiert werden. Da viele dieser Substanzen bioaktiv sind, besitzen sie das Potential, Nicht-Ziel-Organismen in der Umwelt zu schädigen. Einige dieser Substanzen werden kontinuierlich emittiert und sind daher in einem gewissen Ausmass immer präsent in Oberflächengewässern, ein Phänomen welches Pseudopersistenz genannt wird. Die in der Umwelt vorhandene Konzentration einer Chemikalie wird jedoch auch massgebend von ihrer tatsächlichen Persistenz bestimmt, das heisst von der Geschwindigkeit, mit der sie durch biologische und chemische Abbauprozesse entfernt wird. In Oberflächengewässern umfassen die wichtigsten Transformationsprozesse chemische Hydrolyse, direkte und indirekte Phototransformation und mikrobielle Biotransformation. Die Geschwindigkeit und das Ausmaß dieser Transformationsprozesse spielt daher eine wichtige Rolle bei der regulatorischen Risikobewertung von Chemikalien. Für die Gefahrenbeurteilung werden Abbauhalbwertszeiten mit regulatorischen Persistenzkriterien verglichen. Des Weiteren werden für die Expositionsabschätzung Abbauhalbwertszeiten als Eingangsgrößen für Expositionsmodelle verwendet, wie zum Beispiel in den FOCUS-Modellen oder EUSES.

Aufgrund ihrer erhöhten Reproduzierbarkeit und niedrigeren Kosten im Vergleich zu Abbaustudien, welche direkt im Feld durchgeführt werden, spielen Labor-basierte Testsysteme, auch Simulationsstudien genannt, eine wichtige Rolle bei der Bewertung von Abbaubarkeit für regulatorische Zwecke. Gleichzeitig wird davon ausgegangen, dass sie die reale Umwelt besser repräsentieren als einfache Testverfahren für „Leichte Bioabbaubarkeit“ oder Hydrolyse. Für die Bewertung der mikrobiellen Biotransformation von Chemikalien in Oberflächengewässern sind zwei Prüfrichtlinien der OECD relevant: Die OECD 308-Richtlinie ,“Aerobic and Anaerobic Transformation in Aquatic Sediment Systems”, welche Transformation an der Wasser-Sediment-Schnittstelle bewertet, und in jüngster Zeit auch die OECD 309-Richtlinie " Aerobic mineralization in surface water – Simulation biodegradation test ", welche die Bewertung von Transformation im pelagischen Wasserkörper (mit und ohne Schwebstoffe) umfasst. Typischerweise werden drei Arten von Abbauindikatoren aus den Ergebnissen dieser Studien abgeleitet: $DissT_{50,w}$, $DissT_{50,sed}$ und $DegT_{50,ts}$. Diese Parameter beschreiben die Zeit, bis 50% der Ausgangsverbindung aus der Wasserphase, der Sedimentphase oder dem kombinierten System verschwunden sind. Im Laufe der Jahre wurden verschiedene Problemen der OECD 308-Richtlinie erkannt und diskutiert. Hauptkritikpunkte waren die fragliche Relevanz der Testbedingungen für tatsächliche Oberflächengewässer, insbesondere das niedrige Wasser-zu-Sediment-Verhältnis, die geringe Tiefe der Wassersäule, und die stehende Bedingungen. Desweitern wurde die mangelnden Richtlinien und praktischen Hilfsmitteln für die Herleitung von tatsächlichen Abbauhalbwertzeiten aus OECD 308-Daten moniert. Das letztere Problem ergibt sich aus der Tatsache, dass die Konzentrationsmessungen in OECD 308 sowohl Phasentransfer als auch phasenspezifische Transformationsprozesse reflektieren, und der kompartimentsspezifische Abbau deshalb nicht direkt aus den gemessenen Daten ersichtlich ist.

Dieses Forschungsprojekt zielte darauf ab, beide der oben genannten Bedenken aufzugreifen und Empfehlungen abzuleiten, wie bestehenden OECD 308- und OECD 309-Daten am besten zu nutzen sind, um Abbau in Oberflächengewässern zu charakterisieren. Weiter umfasst die Studie Vorschläge, wie man möglicherweise diese Simulationsstudien in Zukunft verbessern oder ergänzen könnte.

In Arbeitspaket I des Projekts wurden sowohl Standard-Ansätze als auch anspruchsvollere inverse Modellierungsansätze angewendet, um unterschiedliche Persistenzindikatoren und ihre Unsicherheitsbereiche aus OECD 308- und OECD 309-Daten für Pestizide und Arzneimittel abzuleiten. Die so hergeleiteten Persistenzindikatoren wurden untereinander und gegen bestehende regulatorische Persistenzkriterien verglichen. Auf dieser Basis wurden Empfehlungen abgeleitet und eine für die Datenauswertung leicht anwendbare Software für die Berechnung der Persistenzindikatoren bereitgestellt. Da alle in diesem Projekt verwendeten OECD 308 und OECD 309-Daten vertraulich waren, wurde im gesamten Bericht ein Code verwendet, um die Substanzen zu identifizieren.

Arbeitspaket II des Projekts behandelt die Frage, inwiefern Labor-basierte OECD 308 und 309 Simulationsstudien das Schicksal von Chemikalien in tatsächlichen Oberflächengewässern wiedergeben können. Zu diesem Zweck wurden in einer Literaturstudie die wichtigsten Einflussgrößen, die den Abbau von Chemikalien an der Wasser-Sediment-Grenzschicht beeinflussen, diskutiert und eine Übersicht erstellt, wie diese in verschiedenen Labor-basierten Testsystemen ausgeprägt sind. Dies beinhaltete auch einen Vergleich von in der wissenschaftlichen Literatur rapportieren Halbwertszeiten in Wasser-Sediment-Systemen zwischen verschiedenen Labor-basierten Systemen und zwischen Labor-basierten Systemen und tatsächlichen Feldstudien. Diese theoretischen Überlegungen wurden mit einer Fallstudie über Chemikalienabbau im Rhein ergänzt. In dieser Fallstudie wurden experimentell bestimmte Halbwertszeiten aus OECD 308-Studien als Eingangsgrößen in ein georeferenziertes Modell des Rheins genutzt, und der Frage nachgegangen, inwiefern Modellergebnisse mit gemessenen Daten aus der Rhein-Überwachung übereinstimmen. Im Folgenden werden die wichtigsten Methoden und Erkenntnisse aus den beiden Arbeitspaketen zusammengefasst.

*Arbeitspaket I: Herleitung von Persistenzindikatoren aus OECD 308- und 309-Daten und Bewertung ihrer Unsicherheit*

Die am direktesten beobachtbare Halbwertszeit, die aus OECD 308 und 309-Daten abgeleiteten werden kann, ist $DegT_{50,ts}$ (häufig auch $DissT_{50,ts}$ genannt). Diese Halbwertszeit beschreibt das Verschwinden der Ausgangsubstanz im gesamten Versuchssystem. $DegT_{50,ts}$ kann direkt aufgrund der beobachteten Abnahme der Ausgangssubstanz im Gesamtsystem mit verschiedenen kinetischen Modellen berechnet werden. Obwohl $DegT_{50,ts}$ häufig als eine „Verschwindens-Halbwertszeit“ interpretiert wird, handelt es sich hierbei auch um eine sinnvolle Abbau-Halbwertszeit, da die Ausgangsverbindung, mit Ausnahme von leichtflüchtigen Verbindungen, nicht aus dem experimentellen Systemen entweichen kann, es sei denn sie wird abgebaut. In Analogie zu $DegT_{50,ts}$ beschreiben $DissT_{50,w}$ und $DissT_{50,sed}$ das Verschwinden aus dem Wasser- und Sedimentkompartiment in OECD 308-Studien. Diese Verschwindens-Halbwertszeiten werden jedoch gleichzeitig durch Biotransformation und Phasentransfer beeinflusst und sind daher nicht geeignet, die Biotransformationseigenschaften der betreffenden Verbindung zu charakterisieren. Kompartiments-spezifische Verschwindens-Halbwertzeiten wurden daher aus unserer Analyse ausgeschlossen. Stattdessen sind Kompartiments-spezifische Abbau-Halbwertszeit im Wasser und Sediment ($DegT_{50,w}$, beziehungsweise $DegT_{50,sed}$) tatsächliche Abbau-Halbwertszeiten und deshalb für die Persistenz- und Expositionsbeurteilung zu bevorzugen. Allerdings müssen sie indirekt aus den Beobachtungen abgeleitet werden, indem ein mechanistisches Modells, welches die Phasenübertragung und Biotransformationswege beschreibt, an die beobachteten Daten angepasst wird.

Im Arbeitspaket I wurde die Methodik zur Ableitung der genannten Persistenzindikatoren aus OECD 308 und 309-Studien entwickelt. Die Methodik wurde in ein Computerprogramm implementiert und für Beispiele von OECD 308 und 309-Datensätzen für 13 Arzneimittel und 14 Pestiziden angewendet. Ein besonderes Augenmerk wurde auf die Definition der Daten und des Datenformates gelegt, welche notwendig sind für eine solche inverse Modellierung des Abbaus und die Unsicherheit der abgeleiteten Indikatoren. Notwendige Daten für die Halbwertszeitberechnungen wurden in einem standardisierten

Datenformat aus allen Experimenten gesammelt, unabhängig davon, ob es sich um OECD 308 oder 309-Daten handelt. Die beiden Elemente dieses Datenformats waren ein Residuen-Tabelle, welche die Dynamik der Kompartiments-spezifischen Radioaktivität der Ausgangsverbindung, aller Transformationsprodukte, $CO_2$ und der nicht extrahierbaren Rückstände aufzeigt, sowie eine zusätzliche Liste von experimentellen Metadaten. Beide experimentellen Systeme, wenn auch in einem geringeren Ausmaß im Falle von OECD 309, sind so komplex, dass es schwierig ist, Biotransformationsprozesse allein auf Basis der beobachteten Konzentrationsdynamik der Ausgangssubstanz und der Transformationsprodukte zu identifizieren. Das Identifikationsproblem bezüglich $DegT_{50,\,w}$ und $DegT_{50,\,sed}$ kann angegangen werden, indem neben der Residuen-Tabelle zusätzliche Metadaten verwendet werden. Diese Metadaten beschreiben die experimentellen Randbedingungen und andere relevante Substanzeigenschaften, die das Verhalten der Verbindung bestimmen. Diese müssen aus unabhängigen Experimenten oder Vorhersagen bekannt sein. Die Zusammenstellen der Daten aus den Prüfberichten gestaltete sich, aufgrund der Vielzahl von Berichtsformaten und Arten von Informationspräsentationen, schwierig. Insbesondere Bestimmungsgrenzen und Substanzverluste in den spezifischen analytischen Methoden, die in Widerspruch zu den Angaben aus der Radioaktivitätsmessung standen, machten es schwierig, eine Residuen-Tabelle mit geschlossener Massenbilanzen zu erstellen. Während fehlende Daten bei der Erstellung von Residuen-Tabellen kein großes Problem waren, wurden gewisse notwendige Metadaten fast nie angegeben. Daten, die häufig fehlten, umfassten die Systemgeometrie, die Porosität des nassen Sediments und der Verteilungskoeffizient zwischen dem organischen Kohlenstoff im Sediment und Wasser ($K_{OC}$) für die Sedimente, welche in dieser Studie verwendet wurden. Um die Probleme der fehlenden Metadaten anzugehen, wurden die fehlenden Eigenschaften teilweise auf der Grundlage der OECD 308-Richtlinie und teilweise aufgrund allgemeinem Wissen abgeschätzt. Messpunkte mit einer schlechten Wiederfindungsrate wurden von der Analyse ausgeschlossen, es sei denn sie leisteten einen entscheidenden Beitrag zur Bestimmung der allgemeinen Form der Zeitreihe. Beispiele für entscheidende Messpunkte umfassen Start- und Endzeitpunkte, sowie zwischenzeitliche Maxima ohne einen benachbarten Messwert von guter Qualität.

$DegT_{50,\,ts}$ wurde durch Anpassung von semi-empirischen kinetischen Modellen berechnet, wie im FOCUS Kinetics Report empfohlen (SFO: single first-order, DFOP: double first-order in parallel, HS: hockey-stick kinetics, FOMC: first-order multi-compartment). Diese Modelle wurden neu implementiert, so dass das gleichen Kalibrierungs- und Unsicherheits-Bewertungsverfahren für $DegT_{50,ts}$ Werte wie für die Kompartiments-spezifischen Halbwertszeiten verwendet werden konnte. Die Berechnung der Kompartiments-spezifischen Biotransformations-Halbwertszeiten wurde mit von (Honti et al. 2016) entwickelten Modellen durchgeführt. Diese Modelle sind in der Lage sowohl OECD 308 als auch OECD 309-Experimente zu simulieren und ermöglichen somit eine Kreuzkalibrierung, das heißt, die Schätzung der Biotransformationsraten auf Basis mehrerer Experimente auf einmal. Im Modell werden die verschiedenen Experimente durch einen oder mehrere voll durchmischte Reaktoren, welche Wasser und Sediment beinhalten, wiedergegeben. Prozesse in den einzelnen Reaktoren wurden basierend auf den folgenden Annahmen modelliert:

- Substanzen sorbieren gemäß ihrer Sorptionseigenschaften ans Sediment. Es wird angenommen, dass der sorbierte Substanzanteil nicht bioverfügbar ist.
- Die Dynamik der Sorption kann durch eine Konvergenz erster Ordnung zum Gleichgewicht hin beschrieben werden.
- Partikuläres organisches Material ist ein guter Indikator für die verfügbare Biomasse unter der Annahme, dass der Anteil aktiver Biomasse am partikulären organischen Material in allen drei Reaktortypen gleich ist. Die Biotransformationsrate ist daher das Produkt einer Geschwindigkeitskonstante ($k'_{bio}$) 2. Ordnung, der Konzentration an partikulärem organischem Kohlenstoff und dem bioverfügbaren Anteil der Substanz.
- Transformationsprodukte werden gesamthaft als ein Substanzpool zusammengefasst.

Modelle für die OECD 308 und 309-Systeme wurden anhand des Zusammenfügens von Reaktoren mit unterschiedlichen Sedimentkonzentrationen erstellt. Beim OECD 309-System handelt es sich um den einfachsten Fall aufgrund des vollständig durchmischten Aufbaus. Pelagische oder nicht-pelagische OECD 309-Experimente können anhand eines einzigen Reaktors mit keinem bzw. wenig gelöstem Sediment modelliert werden. OECD 308-Experimente sind komplexer abzubilden. Das Sedimentkompartiment ist überwiegend anaerob (Honti et al. 2015, Shrestha et al. 2016) mit einer dünnen aeroben Oberflächenschicht, in welcher in der Regel die Mehrheit der Biotransformation stattfindet. Dies bedeutet, dass mehrere Reaktoren für eine realistische Darstellung des Systems zusammengefügt werden mussten. Diese Reaktoreinheiten (d.h. einer für die Wasserphase, einer für die obere aeroben Sedimentschicht, drei anaerobe Sedimentschichten) wurden zwischen benachbarten Elementen mit Diffusionsprozessen verknüpft.

Alle Indikatoren ($DegT_{50,ts}$, $DegT_{50,w}$, $DegT_{50,sed}$) ergeben sich aus der Anpassung des jeweiligen Modells an die beobachteten experimentellen Daten. Aufgrund der Möglichkeit, dass mehrere Parametersätze ähnliche Anpassungsgüte aufweisen, wurde in allen Fällen eine Unsicherheitsbewertung durchgeführt. Dazu wurde eine große Anzahl möglicher Parameterkombinationen ausgewertet. Das Ergebnis der Unsicherheitsbewertung war nicht ein einziger Satz von Modellparametern, welche zu dem niedrigsten möglichen Fehler führen, sondern eine Reihe von Parameterverteilungen, die die statistisch benachbarten Werte der optimalen Werte beschreiben.

(Honti et al. 2016) haben gezeigt, dass die gemeinsame Kalibrierung von verschiedenen Wasser-Sediment-Experimenten für dieselbe Substanz und dasselbe Sediment, die Identifizierbarkeit der Biotransformationseigenschaften der Verbindung erhöhen kann. Verbesserte Identifizierbarkeit und reduzierte Unsicherheit werden erzielt durch die Verwendung von mehr Daten und Daten aus verschiedene experimentellen Systemen, die verschiedene Facetten des Verhaltens der Verbindung in Wasser-Sediment-Systemen darstellen. Während die OECD 309-Richtlinie Biotransformation (oder deren Abwesenheit) in verdünnten, gut gemischten Sedimentsystemen beschreibt, sagt sie nichts über das Verhalten in Gegenwart von geschichtetem Sediment oder unter anaeroben Bedingungen aus. In ähnlicher Weise behandelt die OECD 308-Richtlinie das Verschwinden der Ausgangsverbindung in verschiedenen Kompartimenten, gibt aber keinen zuverlässigen Aufschluss über den Ort und die Mechanismen der Biotransformation. Durch die Behandlung beider Systeme zusammen kann die tatsächliche Biotransformation aus OECD 309-Studien mit suspendiertem Sediment erlernt werden und dadurch die Identifizierbarkeit von Phasentransfer und Biotransformationsprozessen im geschichteten Sediment in der OECD-308 Richtlinie deutlich erhöht werden.

Die Resultate der Modellierung der Daten für Pestizide und Arzneimittel ergaben, dass auch $DegT_{50,ts}$ eine nicht unerhebliche Unsicherheit aufweist. In den meisten Fällen lag die relative Unsicherheit um den Mittelwert bei über 20%, bei etwa der Hälfte der Fälle sogar über 100%. Dies führt zu einer Inkonsistenz zwischen der Unsicherheit im Persistenzindikator $DegT_{50,ts}$ und den derzeit benutzten fixen Persistenzkriterien. Es wird deshalb empfohlen, dass in Zukunft die Unsicherheitsbewertung für $DegT_{50,ts}$ ein integraler Bestandteil der Persistenzbeurteilung werden soll und Verfahren für die Persistenzbeurteilung so weiter entwickelt werden, dass Unsicherheiten in der Persistenzbeurteilung berücksichtigt werden können. Von den untersuchten kinetischen Modellen war SFO das stabilste Modell, um $DegT_{50,ts}$ zu berechnen, welches eine hinreichend gute Anpassung an die beobachteten Daten in allen außer einem Fall erbrachte. Die komplexeren DFOP, HS, und FOMC-Modelle sind für die meisten Experimente überparametrisiert. Sie waren daher numerisch zu instabil für Situationen mit wenig Daten oder Daten außerhalb des beobachteten Datenbereichs.

Die Herleitung der Biotransformationsrate 2. Ordnung, $k'_{bio}$, erwies sich in allen Fällen als sehr unsicher. Dementsprechend waren daraus abgeleitete Kompartiment-spezifische Abbauhalbwertzeiten ebenfalls sehr unsicher. Die Unsicherheit von $DegT_{50,w}$ und $DegT_{50,sed}$ betrug mindestens 1-2 Größen-

ordnungen und die entsprechenden Werte waren, dem Muster von $k'_{bio}$ folgend, verschieden in unterschiedlichen Sedimenten und unter verschiedenen Redox-Bedingungen. Als Folge der begrenzten Verfügbarkeit abbaubauender Biomasse in der Wasserphase, waren $DegT_{50,w}$-Werte immer bis zu mehreren Größenordnungen höher als $DegT_{50,sed}$-Werte für die gleiche Verbindung. Die Tatsache, dass $DegT_{50,w}$ $DegT_{50,sed}$ immer überstieg, steht in interessantem Kontrast mit den verwendeten Persistenzkriterien zur Identifizierung persistenter Verbindungen in den beiden Kompartimenten (40, bzw. 120 Tage für Wasser und Sediment). Dies erklärt auch die praktische Beobachtung, dass für den gleichen Satz von Verbindungen viel mehr OECD 309-Ergebnisse zu einer Klassifizierung als persistent führten als dies für OECD 308-Ergebnisse der Fall war. Es gibt eine schwache für OECD 308 und 309 unterschiedliche Beziehung zwischen $DegT_{50,ts}$ und Kompartiment-spezifischen Halbwertzeiten. Bei stark sorbierenden Verbindungen (d.h., $K_{oc}$ > 5000 L/kg) finden nahezu alle relevanten Prozesse in der Sedimentphase eines OECD 308-Testsystems statt, da sich die Verbindung schnell und praktisch vollständig aus der Wassersäule in das Sediment verlagert. Dies legt nahe, dass $DegT_{50,ts}$ in solchen Fällen in etwa $DegT_{50,sed}$ entsprechen sollte und somit als Persistenzindikator für das Sediment verwendet werden könnte. Die gemeinsame Kalibrierung von verschiedenen Versuchsarten reduziert die Unsicherheit von $k'_{bio}$, $DegT_{50,w}$ und $DegT_{50,sed}$, aber diese Reduktion der Unsicherheit war aufgrund des begrenzten Informationsgehalt von OECD 309-Daten aus pelagischen Systemen gering.

$DegT_{50,w}$ und $DegT_{50,sed}$ von reichlich vorhandenen und dominanten Transformationsprodukten konnte mit vergleichbarer Genauigkeit extrahiert werden wie für die Ausgangsverbindungverbindung, jedoch waren solche Berechnungen auf eine kleine Gruppe von Verbindungen beschränkt, da sie die Dominanz eines einzigen Transformationsproduktes vorraussetzen.

Die Unsicherheit aller Persistenzindikatoren, die aus dynamischen Konzentrationsreihen abgeleitet werden, zeigt eine schwerwiegende Schwäche im Vorgehen bei der Persistenzbeurteilung. Ein Vergleich der optimalen Werte der Persistenzindikatoren mit fixen Werten für die Persistenzkriterien ignoriert deren erhebliche Unsicherheit. Bisher ist jedoch keine robuste und akzeptierte Methode definiert, welche dieses allgegenwärtige Problem für die Persistenzbeurteilung in einem bedeutsamen Rechtsrahmen adressiert.

Die oben beschriebenen Ergebnisse machen deutlich, dass eine signifikante Verbesserung der Persistenzbeurteilung einer Substanz mit relativ geringem Aufwand erreicht werden könnte. Wir schlagen in diesem Zusammenhang vor, geringfügige Änderungen in der Struktur und dem Inhalt der Berichte, wie auch in der Bewertungsmethode vorzunehmen.

Wir schlagen vor, dass die Berechnung von Unsicherheitsintervallen bei der Herleitung von Persistenzindikatoren obligatorisch werden sollte. Für die Berechnung der Kompartiments-spezifischen Abbauhalbwertszeiten schlagen wir vor, die auf $k'_{bio}$–basierenden Modelle zu verwenden. Die gemeinsame Modellierung der OECD 308 und 309-Systeme kann verwendet werden, um die Unsicherheit der Kompartiments-spezifischen Halbwertzeiten zu reduzieren, wobei gezeigt wurde, dass die Ergebnisse der OECD 309-Experimente gemäß Richtlinie nur wenig Informationen über Biotransformation selbst liefert. Der Unterschied zwischen der pelagischen Version von OECD 309 und der Version mit suspendiertem Sediment war kleiner als die typische Messgenauigkeit.

*Arbeitspaket II - Eignung von regulatorischen Daten (OECD 308 und 309-Daten) zur Bewertung der Persistenz in Flüssen*

Im ersten Teil des Arbeitspaketes II wurde eine Literaturrecherche bezüglich experimenteller Systeme zur Bestimmung von Biotransformation und biologischer Abbaubarkeit von Chemikalien in Wasser-Sediment-Systemen durchgeführt. Die Literaturrecherche konzentrierte sich auf standardisierte Testsysteme, welche vor allem für regulatorische Zwecke verwendet werden. Zusätzlich wurden auch stark abweichende oder ergänzende experimentelle Ansätze aus der Literatur erfasst. Die Literaturrecherche

beschränkte sich auf Studien aus den Jahren 2005-2015, mit Ausnahme von einigen früheren Schlüsselstudien, die auch enthalten sind. Basierend auf der Begutachtung von Testsystemen, wurden die wichtigsten Faktoren, die den Abbau in Wasser-Sediment-Systemen beeinflussen können, abgeleitet. Jeder Faktor wurde dann dahingehend erläutert, wie er gemäß Stand des Wissens Biotransformation und somit Persistenz beeinflussen kann, und wo es möglich Bezüge zu gemessenen oder bekannten Charakteristika von Flusssystemen gibt. Schließlich wurden, um die Frage nach der Umweltrepräsentativität der OECD 308 und 309-Testsysteme zu adressieren, vier sehr weit verbreitete, aber deutlich unterschiedliche Gewässertypen in Deutschland im Hinblick auf einige ihrer wichtigsten Merkmale hin beschrieben, die entweder direkt oder indirekt mit den beschriebenen Einflussfaktoren in Zusammenhang stehen.

Unterschiedliche, aber eng verwandte nationale und internationale Prüfrichtlinien wurden gefunden, welche Labor-basierte Tests in einem regulatorischen Kontext beschreiben. Der Hauptunterschied innerhalb dieser Labor-basierten Testsystemen lag darin, ob sie stehend mit geschichtetem Sediment (OECD 308-Typ-Systeme), oder vollständig durchmischt waren (OECD 309-Typ-Systeme). Auch unterschieden sich die Richtlinien im Bezug darauf, ob Licht im Testsystem erlaubt war oder nicht. Des Weiteren gab es Unterschiede darin, wie eng die verschiedenen Einflussfaktoren reguliert wurden. Da einige Prüfrichtlinien eine sehr breite mögliche Realisierung der verschiedenen Einflussfaktoren erlaubten bzw. nicht erlaubten, ist eine große Variabilität in den Ergebnissen der nach unterschiedlichen Prüfrichtlinien durchgeführten Experimente zu erwarten. Um diese Erwartung zu verifizieren, wurden Abbau-Halbwertszeiten für Arzneimittel in Labor-basierten Wasser-Sediment-Systemen, welche in der wissenschaftlichen Literatur zwischen 2005-2015 beschrieben wurden, zusammengestellt. Für Pestizide wurden die zur Verfügung stehenden OECD 308-Daten aus Registrierungsdossiers in Bezug auf die Variabilität der Ergebnisse hin analysiert. Diese begrenzten Daten zeigten, dass die Anwesenheit von Sedimentbiomasse zu einem deutlichen Anstieg der Biotransformationsraten führt. Sobald das Testsystem jedoch tatsächlich Sediment enthält, unabhängig von der Menge, zeigen die Ergebnisse aus verschiedenen Labortestsysteme vernünftige Übereinstimmungen ($DegT_{50,ts}$-Werte innerhalb eines Faktors von vier oder weniger als vier für vier Pharmazeutika). Beim Vergleich von nach OECD 308-Richtlinien durchgeführten Studien waren die experimentellen Ergebnisse noch ähnlicher. Für 17 Pestizide variierten die Verhältnisse zwischen $DegT_{50,ts}$-Werten für zwei verschiedene Sedimente bezüglich der gleichen Verbindung zwischen 1.1 und 2.5 mit einem Medianwert von 1.5. Bei 74 Arzneimittel, für welche OECD 308-Studien mit mehreren Sedimenten verfügbar waren, zeigten nur vier Verbindungen Verhältnisse von $DegT_{50,ts}$ –Werten zwischen zwei Sedimenten von > 10, wobei das Medianverhältnis zwischen $DegT_{50,ts}$ –Werten für zwei verschiedene Sedimente und die gleiche Verbindung 2.0 betrug. Diese Ergebnisse deuten insgesamt darauf hin, dass in Gegenwart von Sedimenten im Dunkeln und innerhalb vernünftiger Variationen im experimentellen Aufbau, ein Unterschied von weniger als einer Größenordnung in $DegT_{50,ts}$ für eine gegebene Verbindung zu erwarten sind.

Die Analyse von vier stark verbreiteten Flusstypen in Deutschland hinsichtlich ihrer Hauptcharakteristika, welche direkt oder indirekt mit den genannten Einflussfaktoren verknüpft sind, zeigte, dass ein bestimmtes Flusssystem unterschiedliche Eigenschaften aufweisen kann, von denen die einen besser durch die Merkmale eines OECD 308-Tests repräsentiert werden und andere wiederum besser durch die Merkmale eines OECD 309-Tests. Es wurde daher gefolgert, dass kein einziges Testsystem allein die Bedingungen in deutschen Flusssystemen im Allgemeinen am besten darstellen könnte, und dass es ebenso wenig ein einzelnes Testsystem gibt, welches die höchste Repräsentativität für alle für die Biotransformation relevanten Charakteristika eines bestimmten Flusssystems besitzt. Aus ähnlichen Gründen wurde zudem auch gefolgert, dass nicht prognostiziert werden kann, welches Testsystem letztlich zu konservativeren Ergebnissen führen würde. Dies gilt umso mehr, da solche relativen Ergebnisse zwischen Testsystemen potentiell für Chemikalien mit stark abweichenden Eigenschaften verschieden sind (z. B. für stark sorbierende Chemikalien oder für Chemikalien, die vorzugsweise unter anaeroben Bedingungen transformiert werden).

Zusammenfassend belegt die theoretische Bewertung der verschiedenen Testsysteme und ihre divergierenden Eigenschaften in Bezug auf Faktoren, von denen erwartet werden kann, dass sie die Biotransformation beeinflussen können, dass Variabilität zwischen den Ergebnissen der Testsysteme unvermeidbar ist. Es wird daher dringend empfohlen, dass im regulatorischen Prozess explizite und transparente Strategien erforderlich sind, um mit diesen umzugehen. Drei verschiedene Strategien wurden dazu anhand der theoretischen Überlegungen und Ergebnisse des Arbeitspakets I skizziert: (i) Auswahl des geeigneten Testsystems auf Substanzbasis, so dass es auf der Grundlage von Stoffeigenschaften und Emissionsszenarien am besten für das Kompartiment steht, in dem sich die meiste Substanzmasse befindet. Ein Beispiel hierfür ist eine aktuelle Empfehlung für stark sorbierende Substanzen in den REACH-Richtlinien, nach welcher OECD 308 einem OECD 309 vorzuziehen ist. Für stark sorbierende Substanzen konnten wir zeigen, dass $DegT_{50,ts}$ tatsächlich $DegT_{50,sed}$ sehr ähnlich ist und dass die Halbwertszeit für das Gesamtsystem daher ein guter Indikator für Persistenz im Sediment ist. Für Substanzen mit mittlerem Sorptionsverhalten ist es jedoch schwieriger, das Testsystem und den entsprechenden Persistenzindikator auf der Grundlage dieser Prinzipien zu wählen, da sich ihre Massenverteilung in Abhängigkeit von der Wassersäulenhöhe und den Sediment-Eigenschaften signifikant zwischen der Sediment- und Wassersäule verschieben kann; (ii) Extrahieren von grundlegenden Informationen über die Biotransformation, welche weniger von der tatsächlichen Testsystemgeometrie und dem Wasser-Sedimentverhältnis eines gegebenen Testsystems abhängen, z. B. die Ableitung einer biomasse- und bioverfügbarkeits-normalisierten Biotransformations-Geschwindigkeitskonstante $k'_{bio}$, wie vorgeschlagen im Arbeitspaket I; (iii) Verwendung des Ergebnisses eines Testsystems, das entweder einem OECD 308, einem OECD 309 entspricht oder einer Varianter davon für die Persistenzbeurteilung, und Anerkennung der damit verbundene Unsicherheit durch Anwenden eines "Sicherheitsfaktors", entweder durch Anpassung des Persistenzkriteriums oder durch Multiplikation des Testergebnisses vor dem Vergleich mit dem Persistenzkriterium.

Im zweiten Teil des Arbeitspakets II wurde die Frage angegangen, wie vergleichbar die in regulatorischen Tests gemessenen Halbwertszeiten mit in tatsächlichen Oberflächengewässern beobachteten Abbauraten sind. Diese Frage wurde auf zwei Arten untersucht: Erstens wurden in der Literatur rapportiere Halbwertszeiten aus Laboruntersuchungen und Feldstudien für ausgewählte Substanzen verglichen. Zweitens wurde eine Feldstudie im Rhein, in welcher das Schicksal diverser Mikroverunreinigungen innerhalb eines Wasserpaketes im Rhein verfolgt wurde, als Fallstudie genutzt. Für die Abschätzung der Halbwertszeiten im Rhein wurden die gemessenen Konzentrationen in diesem sich den Rhein flussabwärts bewegendem Wasserpaket mit Modellvorhersagen verglichen, um die so eruierte Abbaurate im Rhein wiederum mit den aus OECD-308-Daten abgeleiteten Halbwertszeiten zu vergleichen.

In der wissenschaftlichen Literatur wurden zehn Verbindungen gefunden, für welche mindestens eine gemessene Halbwertszeit sowohl in einem Labor-Testsystem als auch aus einer Feldstudie in einem realen Fluss- oder Seesystem zur Verfügung stand. Für diese begrenzte Menge an verfügbaren Daten variierte das Verhältnis zwischen den in dem Feld beobachteten Halbwertszeiten und jenen, die in Labortestsystemen gemessen wurden, nur zwischen 0.19 und 8.4 und war somit für keine Substanz grösser als ein Faktor 10. Es war kein klarer Trend hin zu niedrigeren Verhältnissen von Feld- zu Labor-Halbwertszeiten für Verbindungen mit hohem photochemischem Abbaupotential erkennbar.

In der Rhein-Fallstudie wurde ein georeferenziertes Modell des Rheineinzugsgebietes entwickelt, das hydrologische Informationen für das Jahr 2011, in welchem die Feldstudie mit dem Wasserpaket durchgeführt worden war, beinhaltete sowie Daten zu Standorten und Personenäquivalenten von 2647 Kläranlagen im Einzugsgebiet. Diese Information wurde in AQUASIM, einer Modellierungssoftware für aquatische Systeme, implementiert und mit einem Modell für den Verbleib von Chemikalien in der Wasser-Sediment-Grenzschicht ergänzt. Das Modell wurde verwendet, um Konzentrationsmuster von insgesamt sieben Substanzen vorherzusagen, welche auch in der Feldstudie analysiert wurden.

Dazu gehörten eine konservative Referenzsubstanz (Carbamazepin), vier Substanzen für welche OECD-308-Daten verfügbar waren und zwei Substanzen, deren Konzentrationsprofile einen erheblichen Abbau im Rhein vermuten ließen. Für die Substanzen mit OECD-308-Daten wurden Abbau-Halbwertszeiten, welche aus dem Arbeitspaket I stammten, im Modell implementiert, um die Konsistenz der Verwendung dieser Halbwertszeiten mit den beobachteten Konzentrationen im Feld zu bewerten. Für die beiden Stoffe, welche aufgrund ihrer Konzentrationsmuster als hochgradig abbaubar angesehen wurden, wurden die Halbwertszeiten aufgrund eines Vergleichs von vorhergesagten und gemessenen Konzentrationen abgeschätzt und weiter mit der durchschnittlichen Verweilzeit aller Kläranlagenausflüsse im Rheineinzugsgebiet verglichen. Ergebnisse für die konservative Bezugssubstanz Carbamazepin wurden verwendet, um eine Methode zur Korrektur der vorhergesagten Frachten auf Grundlage der gemessenen Konzentrationen in den wichtigsten Nebenflüssen zu kalibrieren.

Für die vier Substanzen mit OECD 308-Daten wurde für Modellvorhersagen mit Kompartiments-spezifischen Abbauhalbwertszeiten ($DegT_{50,w}$, $DegT_{50,sed}$) kein Widerspruch mit den gemessenen Konzentrationen im Rhein festgestellt. Die Vorhersagen waren jedoch auch nicht deutlich verschieden von (und damit auch nicht eindeutig besser als) Simulationen ohne Abbau. Dies könnte durch die hohen geschätzten Halbwertszeiten im Wasser für alle diese Substanzen erklärt werden. Es wurde festgestellt, dass die Anwendung der Halbwertszeiten für das Totalsystem ($DegT_{50,ts}$) auf das gesamte System, also sowohl im Wasser- als auch im Sedimentkompartiment des Rheinmodells, den Abbau deutlich überschätzte. Dies zeigte, dass $DegT_{50,ts}$ direkt abgeleitet aus dem OECD 308-Test nicht auf Feldsituationen mit anderen Wasser-Sedimentverhältnissen übertragbar ist. Bei zwei Stoffen konnten die anfänglich vorhergesagten Konzentrationsmuster auch unter Annahme verschiedener Abbaukonstanten nicht mit den beobachteten räumlichen Konzentrationsmustern in Einklang gebracht werden. In beiden Fällen zeigten die deutschen und die schweizerischen Verbrauchsdaten große Diskrepanzen auf. Die Ergebnisse verbesserten sich, als der deutsche Verbrauch durch die schweizerischen Verbrauchsdaten ersetzt wurde. Die Unsicherheit über diese Art der Korrektur blieb jedoch bestehen. Diese Beispiele zeigten, dass die Unsicherheit über die räumliche Verteilung des Substanzeintrages ins System die Abschätzung der Abbau-Halbwertszeiten im Rhein stark beeinträchtigen kann und damit auch die Bewertung der Angemessenheit der OECD 308-Daten zur korrekten Vorhersage der Persistenz im Rhein erschwert.

Das Rheinmodell wurde ferner verwendet, um mehrere Abbauszenarien zu vergleichen (d.h. Abbau sowohl in Wasser als auch im Sediment oder Abbau in nur einem Kompartiment). Basierend auf diesen Simulationen wurde klar, dass für die relativ wasserlöslichen untersuchten Substanzen der Abbau in der Wassersäule den beobachteten Gesamtabbau im Fluss dominiert. Dies wurde darauf zurückgeführt, dass der Rhein ein großer Fluss mit einer durchschnittlichen Wassertiefe von etwa 2.4 Metern ist. Diese Ergebnisse unterstrichen, dass die Abbauhalbwertszeiten bezogen auf das System als Ganzes aus einem OECD 308-Experiment, dessen Wasser-Sediment-Verhältnis von 3-4 : 1 etwa um den Faktor 50 niedriger als im Rhein ist, nicht direkt auf einen großen Fluss wie den Rhein übertragen werden kann.

Schließlich wurde das Rheinmodell verwendet, um zu untersuchen, welche Art von Wasserhalbwertszeiten zu einem beobachtbaren Abbau im Rhein führen würden. Es wurde festgestellt, dass Substanzen mit einer Halbwertszeit im Bereich von <6-29 Tagen räumliche Konzentrationsmuster zeigen, die sich deutlich unterscheiden von denen einer konservativen Bezugssubstanz, selbst bei Messunsicherheit und Unsicherheit innerhalb der Modellvorhersagen. Dies steht in Einklang mit der durchschnittlichen Aufenthaltszeit von Substanzen ausgehend von Kläranlagen-Einspeisungen entlang des Rheins, die mittels des georeferenzierten Rheinmodells auf ca. 7.7 Tage geschätzt wurde. Nur Verbindungen, die in der Wassersäule mit einer Halbwertszeit im Bereich der durchschnittlichen Aufenthaltszeit aller Kläranlagenfrachten (d.h. 7.7 Tage) abgebaut werden, weisen einen deutlich erkennbaren Abbau auf.

Wie in Arbeitspaket I gezeigt, werden für die meisten Pestizide und Arzneimittel solche Halbwertszeiten aufgrund der Biotransformation in der Wassersäule nur selten erreicht. Es wurde daher gefolgert, dass nur Substanzen, die eine nennenswerte abiotische Hydrolyse oder Photodegradation zeigen, einen signifikanten Abbau innerhalb des Rheins erwarten lassen. Infolgedessen sind Daten, die aus OECD 308-Studien abgeleitet werden und somit hauptsächlich die Biotransformation im Sediment darstellen, nicht sehr aussagekräftig für den Abbau von Chemikalien in großen Flüssen. Die Ergebnisse abiotischer Hydrolyseuntersuchungen, Photodegradationsstudien oder OECD 309 wären erheblich informativer, um den Abbau von Chemikalien in großen Flüssen darzustellen. Für kleinere Flüsse mit signifikant niedrigeren Wasserständen in der Größenordnung von <0.5 m ist zu erwarten, dass im gesamten Flusssystem mehr Abbau zu beobachten ist und die OECD 308-Richtlinie diese Situation besser widerspiegeln würde. Dies wird auch durch die Ergebnisse des Vergleichs von Halbwertszeiten aus der Literatur für Labor- und Feldsysteme nahegelegt, welche alle kleinere Flüsse als der Rhein behandelten und für welche sich Labor- und Halbwertszeiten bezogen auf das Feldsystem als Ganzes nicht um mehr als einen Faktor von 10 unterschieden.

Die Rhein-Fallstudie erlaubte es auch, die Frage zu betrachten, was ein gutes System wäre, um den Abbau im Feld genau beobachten zu können. Eine wichtige Erkenntnis war, dass für eine gute Persistenzabschätzung im Feld die Unsicherheit in der Inputfunktion deutlich verringert werden muss. Dies kann erreicht werden, indem die Bemessung der Konzentrationen in Kläranlagenabläufen als integraler Bestandteil in Feldstudien miteinbezogen wird, oder indem Studien in kleinen Flüssen mit nur einer Eintragsquelle von Chemikalien durchgeführt werden. Ein Nachteil von Abbauuntersuchungen in kleinen Flüssen ist jedoch, dass die zurückzulegende Strecke und Verweilzeit typischerweise klein sind und daher nur ein schneller Abbau beobachtbar wäre. Ein alternatives Beobachtungssystem wären flache Seen. Diese haben eine erhöhte Verweilzeit im Vergleich zu fließenden Systemen und oft auch nur eine Eingangsquelle. Es ist jedoch unbekannt, wie gut der in diesen Systemen beobachtete Abbau den Abbau in fließenden Gewässern wiederspiegelt. Schließlich wären, wenn auch unerwünscht, zufällige industrielle Substanzeinträge in große Flüsse, wie dies bei dem Unfall in Schweizerhalle 1986 der Fall war, eine weitere Möglichkeit, den Abbau der freigesetzten Chemikalien zu bestimmen. Heute könnten zeitweise industrielle Eingänge, die Teil des regulären Produktionsprozesses sind, eine ähnliche Gelegenheit bieten.

*Abschließende Empfehlungen und offene Diskussionspunkte*

Insgesamt wird auf der Grundlage der Ergebnisse des Projekts die Durchführung von zwei Simulationsstudien zur Beurteilung der Biotransformation in Wasser-Sediment-Systemen empfohlen. Dabei sollte es sich um eine OECD 308-Studie und eine 309-Studie mit so viel suspendiertem Sediment wie erlaubt handeln. Neben der Bereitstellung aller Informationen aus OECD 308-Studien, die für den Beurteilungsprozess notwendig sind, ermöglicht dies die Ableitung von $k'_{bio}$ als fundamentalen Indikator für das Biotransformationspotential einer bestimmten Substanz, welcher Abbauhalbwertzeiten von vielen testsystembezogenen Unterschieden befreit. Dies ermöglicht es, Kompartiments-spezifische Perisistenzindikatoren mit verringerter Unsicherheit zu bestimmen und ermöglicht die Berücksichtigung der tatsächlichen Systemdimensionen im Feld während der Expositionsmodellierung, wie dies in der Rhein-Fallstudie gezeigt wurde.

Ferner wurden durch die Ergebnisse dieses Projekts eine Reihe von Punkten angesprochen, die jedoch nicht oder nur teilweise angegangen werden konnten. Diese stellen Möglichkeiten für weiterführende Forschung dar:

Die Ergebnisse der inversen Modellierung der OECD 308- und OECD 309-Daten (pelagisch) zeigten, dass die Halbwertszeiten in der Wassersäule der 308-Studien und in den OECD 309-Studien hoch wa-

ren, was dazu führte, dass das Persistenzkriterium im Wasser für drei Viertel aller untersuchten Pestizide und Arzneimittel überschritten wurden. Die gemeinsame Modellierung ergab ferner, dass OECD 309-Daten (pelagisch) wenig Informationen über das Ausmaß der Biotransformation lieferten. Dies bedeutet, dass die Durchführung der OECD 309-Richtlinie mit so viel suspendiertem Sediment wie möglich zu einem größeren Informationsgewinn im Bezug auf die Biotransformation in Wasser-Sediment-Systemen führen würde. Die Adaption einer modifizierten OECD 309-Richtlinie mit höheren (1: 100) Sediment-Wasser-Verhältnissen wie bei (Shrestha et al. 2016) würde die Robustheit von $k'_{bio}$ und daraus abgeleiteten Persistenzindikatoren weiter erhöhen. Es müsste jedoch klar kommuniziert werden, dass die direkte Schätzung von $DegT_{50,ts}$ ausgehend von einem solchen System für einen Vergleich mit dem Persistenzkriterium für Wasser wegen der "unnatürlich" hohen suspendierten Sedimentkonzentrationen wahrscheinlich nicht angemessen wäre.

In der Rhein-Fallstudie waren OECD 308-Daten nur für vier der überwachten Stoffe verfügbar, diese waren alle recht polar und wenig abbaubar in der Wassersäule. Es wäre interessant, den Vergleich von gemessenen und vorhergesagten Konzentrationen im Rhein auf Substanzen auszuweiten, welche stärker sorbieren und/oder leichter in der Wassersäule abbaubar sind. Auf diese Weise könnten die Befunde auf einen breiteren Bereich von Stoffverhalten erweitert werden.

Selbst große Flusssysteme wie der Rhein verfügen über maximale Aufenthaltszeiten von Abwasserpaketen in der Größenordnung von nur zwei bis drei Wochen. Auch Substanzen, die eine Persistenz im Wasser von < 40 d aufweisen, würden somit weitgehend ins Meer transportiert. Dies wirft die Frage auf, was das Persistenzkriterium im Wasser eigentlich schützen soll und ob es niedrig genug ist, um die aquatischen Ressourcen ausreichend zu schützen.

In jüngster Zeit wurde die Verwendung eines Benchmark-Ansatzes auf der Basis einer Reihe von Referenzverbindungen mit bekanntem Abbauverhalten als geeignetere Methode zur Beurteilung der Persistenz von Chemikalien vorgeschlagen. Die Verwendung dieser Chemikalien als Bezugssystem und die Bewertung des Verhaltens einer beliebigen Chemikalie gegenüber diesen Referenzchemikalien (und nicht gegen feste Grenzwerte) würde es ermöglichen, die Notwendigkeit einer Extrapolation vom Labor ins Feld zu umgehen und würde auch eine explizite Behandlung der Unsicherheit ermöglichen. Allerdings muss die Nützlichkeit dieses Konzepts noch gründlich erforscht werden. Insbesondere müsste eine Reihe von Referenzchemikalien mit unterschiedlichem Sorptions- und Biotransformationsverhalten definiert werden. Des Weiteren müssten Persistenzmessungen für diese Chemikalien in Labor-basierten Testsystemen und in Feldsystemen durchgeführt werden. Dies würde es ermöglichen, die Frage zu beantworten, ob das relative Verhalten von Substanzen in diesen verschiedenen Systemen ausreichend konserviert ist, um als Referenzsystem für persistente und nicht persistente Chemikalien in Wasser-Sediment-Systemen zu dienen.

# Summary

*Introduction*

For many classes of chemicals a certain loss to surface water bodies over the course of their life-cycle is unavoidable. These include plant protection products (PPPs; i.e., pesticides), which are lost to surface waters through spray drift, run-off or leaching into drainage systems, human pharmaceuticals, biocides and industrial chemicals, which are only partially removed during wastewater treatment and thus are emitted to receiving waters, and finally also veterinary medicines, which through application of manure or outdoor husbandry may reach soil and from there again be lost to surface waters. Since many of these substances intentionally exhibit biological activity, they bear the potential to harm non-target organisms in the environment. Some of them are continuously emitted and therefore are always present in surface water bodies to some extent, a phenomenon termed pseudo-persistence. However, the actual levels at which a chemical is observed and for how long it remains in the environment after emission has ceased is determined by its persistence, i.e., by the rate at which it is removed by biological and chemical degradation processes. For surface water systems, the most important transformation processes include chemical hydrolysis, direct and indirect phototransformation and microbial biotransformation. The speed and extent of these transformation processes therefore plays an important role in the regulatory risk assessment of chemicals. For hazard assessment, transformation half-lives are compared to regulatory persistence criteria, and, for exposure assessment, transformation half-lives are used as inputs for exposure models such as the FOCUS models or EUSES.

Laboratory-based test systems, also called simulation tests, play an important role in evaluating chemical transformation for regulatory purposes due to their increased reproducibility and lower costs when compared to tests conducted in the field. Yet, compared to lower-tier biodegradability and hydrolysis tests, they are closer to representing a real environmental situation. For the evaluation of the microbial biotransformation of chemicals in surface water systems, two OECD testing guidelines are relevant: The OECD 308 guideline on "Aerobic and Anaerobic Transformation in Aquatic Sediment Systems", which assesses transformation at the water-sediment interface, and more recently also the OECD 309 guideline "Aerobic mineralization in surface water – Simulation biodegradation test", which assesses transformation in the pelagic water body (with and without suspended sediment). Typically, three types of degradation indicators are derived from the outcomes of these tests, namely $DissT_{50,w}$ , $DissT_{50,sed}$ and $DegT_{50,ts}$. These parameters describe the time until 50 % of the parent chemical has disappeared from the water phase, the sediment phase or the combined system, respectively. Over the years, different issues with OECD 308 have been reported and discussed. Main points of criticism were concerns about the relevance of the test conditions to properly reflect degradation in actual surface water bodies (i.e., low water-sediment ratio, shallow depth of water column, stagnant conditions), and a lack of guidance and practical tools for the derivation of half-lives from OECD 308 data for use in exposure modeling and persistence assessment. The latter problem stems from the fact that the measured concentration dynamics in OECD 308 reflect both phase transfer and phase-specific transformation processes and that degradation is therefore not directly observable from the measured data.

This research project aimed at addressing both of the above-mentioned concerns to derive recommendations as to how to best make use of existing OECD 308 and OECD 309 data to characterize degradation in actual surface waters, and also how to potentially improve or complement these simulation tests in the future. Specifically, in work package I of the project, both standard approaches and more sophisticated inverse modeling approaches were applied to derive different persistence indicators and their uncertainty ranges from OECD 308 and OECD 309 data for pesticides and pharmaceuticals. The different persistence indicators thus obtained were compared against each other and against existing regulatory persistence criteria. Based on this, recommendations for future data evaluation were derived and readily applicable software for the calculation of persistence indicators is provided. Since all

OECD 308 and OECD 309 data used in this project were confidential data provided by the German Environment Agency, anonymized codes are used throughout the report to identify those substances. Work package II of the project addressed the representativeness of the laboratory-based OECD 308 and 309 simulation tests to reflect and predict the chemicals' fate in actual surface water bodies. For this purpose, a literature review discussing the major factors influencing chemical degradation in surface water bodies and how these are reflected in different laboratory-based test systems was conducted. This also included a comparison of half-life data reported in the scientific literature for different laboratory-based systems and between laboratory-based systems and actual field situations. These theoretical considerations were complemented with a case study on substance degradation in the river Rhine. In this case study, monitoring data from the river Rhine were combined with experimental data from OECD 308 studies and a chemical fate model to evaluate the appropriateness of OECD 308 data to reflect degradation in a large stream such as the Rhine. In the following, the main methodologies and findings from the two work packages are summarized.

*Work package I: Derivation of persistence indicators and their uncertainties from OECD 308 and 309 data*

The most directly observable half-life to be derived from OECD 308 and 309 data is $DegT_{50,ts}$, that is the disappearance half-life of the parent compound from the entire experimental system. This can be computed right from the observed overall decline of the parent compound across all compartments with different kinetic models. Despite being often interpreted as a disappearance half-life, $DegT_{50,ts}$ is a meaningful degradation half-life due to the fact that the parent compound – except for highly volatile compounds – cannot escape the experimental systems unless transformed. In analogy to $DegT_{50,ts}$, $DissT_{50,w}$ and $DissT_{50,sed}$ describe disappearance from the water and sediment compartments in OECD 308. However, these are known to merge biotransformation and phase transfer and are therefore not suitable to characterize the biotransformation properties of the subject compound. Dissipation half-lives were therefore excluded from our analysis. Instead, compartment-specific degradation half-lives in water and sediment ($DegT_{50,w}$ and $DegT_{50,sed}$, respectively) are true degradation half-lives and should therefore be preferred for persistence and exposure assessment. However, they need to be derived indirectly from the observations by fitting a mechanistic model to the data that describes both phase transfer and biotransformation pathways.

In work package I, the methodology to derive these degradation half-life indicators from OECD 308 and 309 studies was developed, encoded into a computer program, and applied to the example of OECD 308 and 309 datasets for 13 pharmaceuticals and 14 pesticides. Particular emphasis was placed on defining the data and data format needed for such inverse modelling of degradation half-lives to be meaningful, and the uncertainty of the derived indicators. Necessary data for half-life calculations were gathered in a standardised data format from all experiments, regardless of experimental type. The two elements of this data format were a residue table describing the time-dynamics of compartment-specific radioactivity of the parent compound, all transformation products pooled together, $CO_2$, non-extractable residues, and an additional list of experimental metadata. Modeling OECD 308 or 309 systems is difficult based on a residue table alone. Both experimental systems, although 309 to a lesser degree, are so complex that biotransformation processes are difficult to identify just based on the observed time-dynamics of the parent compound and transformation products. The identification problem affecting $DegT_{50,w}$ and $DegT_{50,sed}$ can be tackled by utilising additional metadata besides the residue table. These metadata describe the experimental boundary conditions and other relevant environmental fate properties of the compound that need to be known from independent experiments or prediction.

Data collection faced numerous difficulties due to the variety of report formats and kinds of information presented. Many reports followed a structure that mirrored measurement stages exactly, which

made it difficult to assemble a residue table with closed mass balance. While missing data were of less concern for residue tables, certain required metadata were almost never reported. These included system geometry, the porosity of wet sediment, and the organic carbon-water partition coefficient ($K_{oc}$) for the exact sediment used in the studies. To overcome the problems of missing metadata, we relied on expert estimates partly based on the OECD 308 standard and partly on common knowledge for the non-reported properties. Measurement records with poor recovery were excluded from analysis, unless the specific measurement record was a crucial point determining the general shape of the time series. Examples of crucial points include starting, central, or final time points without a close neighbouring record of good quality.

Total system half-life was calculated by fitting the semi-empirical kinetic models recommended by the FOCUS Kinetics Report (SFO: single first-order, DFOP: double first-order in parallel, HS: hockey-stick kinetics, FOMC: first-order multi-compartment) to the overall amount of the parent compound in the entire experimental system. These models were re-implemented by us so that the same calibration and uncertainty assessment procedure used for compartment-specific half-lives can also be carried for $DegT_{50,ts}$.

The calculation of compartment-specific biotransformation half-lives was carried out within the model framework developed by Honti et al. (2016). This model framework is capable of simulating both OECD 308 and OECD 309 experiments and therefore allows cross-experiment calibration, i.e., the estimation of biotransformation rates from multiple experiments at once. The basic building block of the model framework was a fully-mixed reactor layer containing both water and sediment. Processes in the reactor were modelled based on the following assumptions:

- Sorption takes place to sediment. Sorbed fractions are not considered bioavailable.
- The dynamics of sorption can be described by a first-order convergence towards equilibrium.
- Particulate organic matter is a good proxy for the available biomass, assuming that the fraction of active degraders within the particulate organic matter is the same in all three reactor types. Biotransformation rate is therefore a product of a second-order rate constant ($k'_{bio}$), the particulate organic carbon concentration and the proportion of bioavailable fractions.
- Transformation products are merged into a unified pool.

Models for OECD 308 and 309 systems were built by assembling reactors of different sediment concentration. OECD 309 systems are the simplest case due to their fully mixed nature. Pelagic or non-pelagic OECD 309 can be modelled by taking a single reactor with none or some suspended sediment, respectively. OECD 308 experiments are more complex. The sediment compartment is predominantly anaerobic (Honti et al. 2015, Shrestha et al. 2016) with a thin aerobic surface layer where the majority of biotransformation usually takes place. This meant that multiple stacked reactors were necessary for a realistic representation of the system. These reactor layers (1 for water phase, 1 aerobic settled sediment, 3 anaerobic settled sediments) were linked by diffusion processes between adjacent elements.

All indicators ($DegT_{50,ts}$, $DegT_{50,w}$, $DegT_{50,sed}$) derive from fitting a model to the observed experimental data. Due to the possibility of multiple parameter sets having similar goodness of fit, we performed an uncertainty assessment in all cases. A high variety of parameter combinations were evaluated, the result of the uncertainty assessment was not a single set of model parameters leading to the lowest possible error but rather a set of parameter distributions that statistically describe the adjacent values of the optimal values.

It was demonstrated by Honti et al. (2016) that the joint calibration of different water-sediment experiments belonging to the same compound and sediment could increase the identifiability of biotransformation properties of the compound. Improved identifiability and reduced uncertainty is due to the involvement of more data and more aspects. Different experimental standards grasp different facets of

the compound's behavior in water-sediment systems. While OECD 309 describes biotransformation (or its absence) in thin, well-mixed systems, it does not tell anything about behavior in the presence of settled sediment or under anaerobic conditions. Similarly, OECD 308 tells about the disappearance of the parent compound in different compartments, but it does not give solid evidence about the place and mechanisms of biotransformation. Thus, by treating the two types together, actual biotransformation in water can be learnt from pelagic OECD 309 data and thereby the identifiability of phase-transfer and sediment biotransformation processes in the OECD 308 may increase significantly.

The modelling of the pesticide and pharmaceutical data showed that $DegT_{50,ts}$ also had a significant uncertainty. For most cases the relative uncertainty around the mean exceeded 20%, for about half of the cases it exceeded 100%. This leads to an incompatibility of the uncertain $DegT_{50,ts}$ and the presently used rigid persistence criteria. In the future, uncertainty assessment for $DegT_{50,ts}$ should become an integral part of persistence assessment and persistence classification procedures should be further developed to be able to account for this uncertainty in persistence assessment. From the investigated kinetic models, SFO was the most robust way to calculate $DegT_{50,ts}$, giving sufficiently good fit to the observed data in all but one case. The more complex DFOP, HS, and FOMC models are over-parameterised for most experiments. They were therefore numerically unstable for scarce data or outside the observed data range.

The second-order biotransformation rate constant $k'_{bio}$ turned out to be very uncertain in all cases. Accordingly, compartment-specific degradation half-lives were uncertain too. The uncertainty of $k'_{bio}$ mapped into uncertainty in $DegT_{50,w}$ and $DegT_{50,sed}$. Both $DegT_{50,w}$ and $DegT_{50,sed}$ were uncertain up to at least 1-2 orders of magnitude, and, following the patterns of $k'_{bio}$, they were different in different sediments and under different redox conditions. As a result of the limited availability of degrader biomass in the water phase, $DegT_{50,w}$ values were always higher than $DegT_{50,sed}$ values for the same compound by up to several orders of magnitude. The fact that $DegT_{50,w}$ always exceeded $DegT_{50,sed}$ was in interesting contrast with the persistence criteria used to identify persistent compounds in these two compartments (40 and 120 days for water and sediment, respectively). However, it explained the practical experience that for the same set of compounds much more OECD 309 results led to a persistent classification than OECD 308 results did. There are weak relationships between $DegT_{50,ts}$ and compartment-specific half-lives, which differed for OECD 308 and 309. For strongly sorbing compounds (i.e., $K_{oc}$ > 5000 L/kg), almost all relevant processes take place in the sediment phase of an OECD 308 due to the immediate and full migration of the spiked compound from the water column into the sediment. This suggests that $DegT_{50,ts}$ should roughly equal $DegT_{50,sed}$ in such cases and could potentially be used as a persistence indicator. The joint calibration of different experimental types reduced the uncertainty of $k'_{bio}$, $DegT_{50,w}$ and $DegT_{50,sed}$, but this reduction of uncertainty was limited due to the very limited extent of biotransformation observed in OECD 309 systems.

$DegT_{50,w}$ and $DegT_{50,sed}$ of abundant and dominant transformation products could be extracted with comparable accuracy as for the parent compound, yet the required abundance and dominance of a single transformation product limited such exercises to a small set of compounds.

The uncertainty inherent in all persistence indicators derived from concentration-time series data highlighted a serious weakness of the present persistence assessment framework. Comparison of median half-lives values to persistence criteria ignores the significant uncertainty, yet a robust and accepted way of acknowledging this omnipresent uncertainty is not defined in any regulatory framework relevant for persistence assessment. The above outlined issues and results highlight that significant improvement could be made to the assessment of a compounds' degradation process with relatively small efforts. We suggest minor modifications in the structure and the content of the reports, and also in the assessment method.

We suggest to make the calculation of the uncertainty of all reported half-lives obligatory. For the calculation of compartment-specific degradation half-lives, we suggest to use the model framework based

on $k'_{bio}$. The joint modeling of OECD 308 and 309 systems could be used to reduce the uncertainty of the compartment-specific half-lives, yet the standard OECD 309 was found to provide only weak information about biotransformation itself. The difference between the pelagic and suspended sediment versions of OECD 309 was smaller than the typical measurement accuracy.

*Work package II – Suitability of regulatory data (OECD 308 and 309 data) to predict degradation in rivers*

In the first part of work package II, a review of the existing literature on experimental systems to determine biotransformation or biodegradation of chemicals in water-sediment systems was conducted. The review focused on standardized test systems that are mainly used for regulatory purposes, but also reviewed the scientific literature for reports on strongly deviating or complementary experimental approaches. The literature review was restricted to covering studies published in the years 2005-2015 with the exception of some earlier key studies that are also included. Based on the review of test systems, the most important factors influencing degradation in water-sediment systems were derived. Each factor was then described with respect to the state-of-art knowledge on how it may influence persistence and, where possible, it was related to readily measurable or known characteristics of river systems. Finally, to address the question about the environmental representativeness of OECD 308 and 309 test systems, four highly prevalent, yet distinctly different river types in Germany were described in terms of some of their major characteristics that are either directly or indirectly linked to those influencing factors.

Different, yet closely related national and international testing guidelines were found that described laboratory-based tests used in a regulatory context due to their better reproducibility and lower costs compared to field-based experiments conducted in real-world aquatic systems. The main distinction within those laboratory-based test systems was whether they were stagnant, layered systems (i.e., OECD 308-type systems) or fully mixed, suspended sediment systems (i.e., OECD 309-type systems). Also, guidelines differed with respect to whether lighting of the test system was allowed or not, and generally with respect to how narrowly the different influencing factors were regulated. Overall, since some testing guidelines allowed for a very wide possible realization of the different influencing factors or did not regulate them at all, large variability in the outcomes of tests carried out according to different testing guidelines would be expected. To ground-proof this expectation, degradation half-lives for pharmaceuticals with half-lives from laboratory-based testing of water-sediment systems reported in the peer-reviewed literature between 2005-2015 were compiled. For pesticides, the available 308 data from registration dossiers was analyzed with respect to the variability of the results. These limited data showed that the presence of sediment biomass led to a clear increase in the rates of biotransformation. Once the test system included some sediment, however, the outcomes from different laboratory test systems showed reasonable agreement ($DegT_{50,ts}$ values well within a factor of four or less for four pharmaceuticals). When comparing only across OECD 308 studies, experimental results were even more similar. For 17 pesticides, the ratios between $DegT_{50,ts}$ values for two different sediments and the same compound varied between 1.1 and 2.5, with a median value of 1.5. For 74 pharmaceuticals for which OECD 308 studies with multiple sediments were available, only four compounds exhibited ratios of $DegT_{50,ts}$ values between two sediments of >10, and the median ratio between $DegT_{50,ts}$ values for two different sediments and the same compound was 2.0. These results, overall, suggested that in the presence of sediment, in the dark and within reasonable variations in experimental setups, differences of less than an order of magnitude in $DegT_{50,ts}$ values for a given compound can be expected.

The analysis of four highly prevalent river types in Germany with respect to their major characteristics that are either directly or indirectly linked to those factors thought to influence degradation in water-sediment systems showed that a given river system can have different characteristics that are most suitably represented by features of either an OECD 308 or 309 test, and that while one characteristic might be best represented by one of the test systems, another characteristic of the same river might be

more suitably represented by the other test system. It was therefore concluded that no single test system could best represent the conditions in German river systems in general, nor is there even a single test system that is most representative of a given type of river system in all aspects relevant to biotransformation. On similar grounds it was also concluded that it cannot be predicted which test system would ultimately yield more conservative results. This is even more so since such relative outcomes between test systems would potentially be different for chemicals with strongly deviating properties (e.g., for strongly sorbing chemicals or for chemicals that are preferentially transformed under anaerobic conditions).

In conclusion, the theoretical evaluation of the different test systems and their divergent properties with respect to factors that can be expected to influence biotransformation clearly demonstrated that variability between test system outcomes is unavoidable. It is therefore strongly recommended that in the regulatory process explicit and transparent strategies are needed to deal with this. Three different strategies to do so were outlined based on the theoretical considerations and results from work package I: (i) Choosing the appropriate test system on a substance-by-substance basis such that, based on substance properties and emission scenarios, it best represents the exposure compartment where most of the substance mass will reside. An example of this is the current recommendation in the REACH Information Requirements (2016) for strongly sorbing substances to prefer testing according to OECD 308 over OECD 309. For these strongly sorbing substances, we could show that $DegT_{50,ts}$ is actually very similar to $DegT_{50,sed}$, and that the total system half-life is indeed a good indicator of persistence in sediment. However, for substances with intermediate sorption behavior, it is more difficult to choose the test system and appropriate persistence indicator based on these principles because their mass distribution may shift significantly between the sediment and water column, depending on water column height and sediment properties; (ii) Extracting more fundamental information on biotransformation that is less dependent on the actual test system geometry and water:sediment ratio of a given test system, e.g., deriving a biomass- and bioavailability-normalized biotransformation rate constant $k'_{bio}$ as suggested in work package I; (iii) Using the outcome of any test system that is either an OECD 308, an OECD 309 or a variant thereof for persistence assessment, and acknowledge the uncertainty inherent in doing so by introducing a "safety factor", either in the persistence criterion itself or by multiplying the test outcome prior to comparison to the persistence criterion.

In the second part of work package II, the question was addressed how comparable the half-lives measured in regulatory tests are relative to degradation half-lives observed in actual surface water bodies. This question was addressed in two ways: First, literature-reported half-lives were compared for substances for which half-lives were reported from both laboratory studies and field studies. Second, a field study in the river Rhine where the fate of diverse micropollutants in a parcel of water was followed down the Rhine was used as a case study. The measured concentrations in this parcel of water, as it was traveling down the Rhine, were used to estimate half-lives in the Rhine, which could then again be compared to the half-lives derived from OECD 308 data.

As for the literature-reported half-lives, ten compounds were found for which at least one measured half-life in both a laboratory test system and from a field study in a real river or lake system was available. For this limited amount of available data, the ratio between the half-lives observed in the field and those measured in laboratory test systems varied between 0.19 and 8.4 only, and thus did not exceed a factor of 10 for any compound. No clear trend towards lower field-to-laboratory half-life ratios in the case of compounds with known high photochemical degradation could be recognized.

In the Rhine case study, a georeferenced model of the Rhine catchment was developed that included hydrological information for the year 2011 in which the field study on the water parcel had been conducted, as well as the location and person equivalents of 2647 wastewater treatment plants (WWTPs) in the catchment. This information was implemented into AQUASIM, a modeling environment for aquatic systems, and supplemented with a model for the fate of chemicals across the water-sediment

boundary layer. The model was used to predict concentration patterns of altogether seven substances that had also been analyzed in the field study. These included one conservative benchmark substance (i.e., carbamazepine), four substances with available OECD 308 data, and two substances whose concentration profiles suggested considerable degradation in the river Rhine. For the substance with OECD 308 data, degradation half-lives as derived from work package I were implemented into the model runs to evaluate the consistency of using these half-lives with the observed concentrations in the field. For the two substances which were considered highly degradable based on their concentration patterns, half-lives were estimated from a comparison of predicted and measured concentrations and compared to the average travel time of all WWTP inputs in the Rhine catchment. Results for the conservative benchmark chemical carbamazepine were used to calibrate a methodology to correct the predicted input loads based on measured concentrations in the main tributaries.

For the four substances with OECD 308 data available, predictions using compartment-specific degradation half-lives ($DegT_{50,w}$, $DegT_{50,sed}$) were found to not contradict measured concentrations in the Rhine, but to be not much different from (and hence also not superior to) simulations assuming no degradation. This could be explained by the high estimated half-lives in water of all of these substances as estimated from inverse modeling of the OECD 308 data. Instead, it was found that the application of the total system half-life ($DegT_{50,ts}$) as degradation half-life to both the water and sediment compartment of the Rhine model clearly overestimated degradation. This demonstrated that $DegT_{50,ts}$ directly observed in the OECD 308 test system is not transferable to field situations with other water-sediment ratios. There were also two substances for which the initially predicted concentration patterns could not be brought in agreement with observed spatial concentration patterns, even when assuming different extents of degradation. In both cases, German and Swiss consumption data showed large discrepancies and results improved when German consumption was replaced by Swiss consumption data. However, uncertainty about this type of correction remained. These examples showed that uncertainty about the spatial distribution of substance input can severely impede the estimation of degradation half-lives in the river Rhine, and hence also makes assessing the appropriateness of OECD 308 data to correctly reflect substance degradation in the river Rhine difficult.

The river Rhine model was further used to compare several degradation scenarios (i.e., degradation in both water and sediment, or degradation in one compartment exclusively). Based on these simulations, it became clear that for the fairly water-soluble substance studies, degradation in the water column dominates the observed overall degradation in the river. This was attributed to the fact that the river Rhine is a large stream with an average depth of the water column of about 2.4 meters. These findings underlined that total system degradation half-lives obtained from an OECD 308 experiment, whose water-sediment ratio of 3-4:1 is about a factor of 50 lower than in the Rhine, cannot be directly transfer to a large stream like the Rhine.

Finally the river Rhine model was used to investigate what kind of water half-lives would result in an observable degradation along the river Rhine. It was found that substances with half-lives in the range of <6-29 days would show spatial concentration patterns that are clearly different from those of a conservative benchmark chemical, even given measurement uncertainty and uncertainty in model predictions. This is in agreement with the average travel time of all WWTP input loads along the river Rhine, which was derived using the georeferenced Rhine model to be approximately 7.7 days. Only compounds that degrade in the water column with a half-life in the range of the average travel time of all WWTP input loads (i.e., 7.7 days) can be expected to show a clearly recognizable degradation signature. As demonstrated in work package I, for most pesticide and pharmaceuticals, such half-lives are rarely achieved due to biotransformation in the water column. It was therefore concluded that only substances that either show appreciable abiotic hydrolysis or photodegradation are expected to show significant degradation in the river Rhine. As a consequence, data obtained from OECD 308 studies, which are mostly representing biotransformation in the sediment, are not very relevant to representing

degradation of chemicals in large rivers. Rather, results of abiotic hydrolysis studies, photodegradation studies, or OECD 309 would be considerably more informative to represent degradation of chemicals in large rivers. For smaller rivers with significantly lower water levels on the order of < 0.5 m, more degradation can be expected to be observed in the overall river system, and OECD 308 might more appropriately reflect this situation. This is also suggested by the results of the comparison of literature half-lives for laboratory and field systems, all of which were representing smaller rivers than the river Rhine, and for which laboratory and field total system half-lives did not differ by more than a factor of 10.

The Rhine case study also allowed considering the question what would be a good system to accurately observe degradation in the field. The results of the Rhine case study clearly show that uncertainty in the inputs needs to be reduced. Therefore, either monitoring of representative WWTP outflows should be included as integral part into such field studies or studies should be carried out in small streams with only one source of chemicals to the stream. However, one drawback of studying degradation in small rivers is that the travel distances are typically rather small and that only rapid degradation is therefore observable at all. An alternative observation system would be shallow lakes. These have an increased residence time compared to flowing systems and often only one source of input also. However, it is unknown how well degradation observed in those systems represents degradation in flowing rivers. Finally, although unwanted, accidental industrial substance spills into large rivers, such as was the case during the accident in Schweizerhalle in 1986, would be another opportunity to determine degradation of the chemicals released in the spill. Today, temporally intermittent industrial inputs that are part of the regular production process may present a similar opportunity.

*Final recommendations and open discussion points*

Overall, based on the results of the project, the execution of two simulation studies to assess biotransformation in water-sediment systems is recommended. These should be an OECD 308 study and a 309 study with as much suspended sediment as allowed. This would allow, besides supplying all the different pieces of information that OECD 308 studies are required for during the assessment process, derivation of $k'_{bio}$ as a fundamental indicator of the biotransformation potential of a given substance that gets rid of many of the test system-related differences in degradation half-lives. This in turn allows deducing compartment-specific half-life indicators with reduced uncertainty, and allows considering the actual system dimensions in the field during exposure modelling as demonstrated in the Rhine case study.

Further, a number of points were raised by the findings of this project, which could, however, not or only partially be addressed. These present potential opportunities for follow-up research:

The results from inverse modeling of OECD 308 and OECD 309 (pelagic) data indicated that half-lives in the water column of the 308 study and in the OECD 309 study were long, to the effect that three quarters of all pesticides and pharmaceuticals studied exceeded the persistence criterion in water. Joint modeling further indicated that OECD 309 (pelagic) data contributed little information as to the extent of biotransformation. This suggests that running OECD 309 with as much suspended sediment as possible would lead to a larger information gain with respect to biotransformation in water-sediment systems. Adoption of a modified OECD 309 with higher (1:100) sediment-water ratios as was done in Shrestha et al. (2016) would further increase the robustness of $k'_{bio}$ and hence persistence indicators derived therefrom. However, it would need to be understood that direct estimation of $DegT_{50,ts}$ from such a system for comparison to the persistence criteria in water would likely not be appropriate because of the “unnaturally” high suspended sediment concentrations.

In the Rhine case study, OECD 308 data was only available for four of the substances monitored and all of them were fairly polar and recalcitrant in the water column. It would be interesting to extend the

comparison of measured and predicted concentrations in the Rhine to substances that sorb more strongly and/or degrade more readily in the water column. In this way, the findings could be generalized to a broader range of substance behavior.

Even large river systems such as the Rhine have maximal travel times of wastewater packages on the order of three weeks. Substances that pass the persistence criterion in water of < 40 d would thus still be transported to a large extent into the sea. This raises the question what the persistence criterion in water should actually protect and whether it is low enough to sufficiently protect aquatic resources.

More recently, the use of benchmarking based on a set of reference compounds with well-known environmental degradation behavior has been suggested as a more appropriate way to assess persistence of chemicals. Using those chemicals as a reference system and assessing the behavior of any chemical against those reference chemicals (rather than against some fixed persistence criteria) would allow circumventing the need for lab-to-field extrapolation and would also allow for a more explicit treatment of uncertainty. However, the usefulness of this concept still needs to be thoroughly explored. In particular, a set of reference chemicals with diverse sorption and biotransformation behavior would need to be defined and results for these chemicals in laboratory-based test systems and in field systems would need to be obtained. This would allow answering the question whether the relative behavior of substances is sufficiently conserved across these different systems to serve as a reference system for persistent and not persistent chemicals in water-sediment systems.

# 1. Introduction

For many classes of chemicals a certain loss to surface water bodies over the course of their life-cycle is unavoidable. These include plant protection products (i.e., pesticides), which are lost to surface waters through run-off or leaching into drainage systems, human pharmaceuticals, biocides and industrial chemicals, which are only partially removed during wastewater treatment and thus are emitted to receiving waters, and finally also veterinary medicines, which through application of manure or outdoor husbandry may reach soil and from there again be lost to surface waters (Schwarzenbach et al. 2006). More recently, non-target analysis of water quality in larger rivers in Germany and Switzerland has rekindled interest in direct emissions of production intermediates or even end products from chemical and pharmaceutical industry (Ruff et al. 2015, Schlüsener et al. 2015).Many of these substances possess a wanted biological activity, and hence bear the potential to also harm non-target organisms in the environment. Some of them are continuously emitted and therefore are always present in surface water bodies to some extent, a phenomenon termed pseudo-persistence. However, the actual levels at which a chemical is observed and for how long it remains in the environment after emission has ceased is determined by its persistence, i.e., by the rate at which it is removed by biological and chemical degradation processes (Boethling et al. 2009). For surface water systems, the most important transformation processes include chemical hydrolysis, direct and indirect phototransformation and microbial biotransformation. The speed and extent of these transformation processes therefore plays an important role in the regulatory risk assessment of chemicals. For hazard assessment, transformation half-lives are compared to regulatory persistence criteria (ECHA 2008), and, for exposure assessment, transformation half-lives are used as inputs for exposure models such as the FOCUS models or EUSES (FOCUS 1997, VICH VICH00 2004, ECHA 2016c).

Laboratory-based test systems, also termed simulation tests, play an important role in evaluating chemical transformation for regulatory purposes due to their increased replicability and economic nature when compared to tests conducted in the field. Yet, compared to lower tier biodegradability and hydrolysis tests, they are closer to representing a real environmental situation. For the evaluation of the microbial biotransformation of chemicals in surface water systems, two OECD testing guidelines are currently relevant: The OECD 308 guideline on “Aerobic and Anaerobic Transformation in Aquatic Sediment Systems”, which assesses transformation at the water-sediment interface, and more recently also the OECD 309 guideline “Aerobic mineralization in surface water – Simulation biodegradation test”, which assesses transformation in the pelagic water body (with and without suspended sediment). Typically, three types of degradation indicators are derived from the outcomes of these tests, namely $DissT_{50,w}$, $DissT_{50,sed}$ and $DissT_{50,ts}$. These parameters describe the time until 50 % of the parent chemical has disappeared from the water phase, the sediment phase or the combined system, respectively (FOCUS 2006). Over the years, different issues with OECD 308 have been reported and discussed (Bowmer et al. 2004, Davis et al. 2005, Ericson 2007, ECETOC 2010, EMA 2011, ECETOC 2014, Ericson et al. 2014, Radke et al. 2014). Main points of criticism were concerns about the relevance of the test conditions to properly reflect degradation in actual surface water bodies (i.e., low water-sediment ratio, shallow depth of water column, stagnant conditions), and a lack of guidance and practical tools for the derivation of half-lives from OECD 308 data for use in exposure modeling and persistence assessment. The latter problem stems from the fact that the measured concentration data in OECD 308 reflect both phase transfer and phase-specific transformation processes and that degradation is therefore not directly observable from the measured data.

This research project aimed at addressing both of the above-mentioned concerns to derive recommendations as to how to best make use of existing OECD 308 and OECD 309 data, and also how to potentially improve or complement these simulation tests in the future. Specifically, in Chapter 2, both standard approaches (FOCUS 2006) and more sophisticated inverse modeling approaches (Honti et al. 2015) are applied to derive different persistence indicators and their uncertainty ranges from OECD

308 and OECD 309 data for pesticides and pharmaceuticals. The different persistence indicators, both traditional and derived from inverse modeling, are compared against each other and against existing persistence criteria. Based on this, recommendations for future data evaluation and a readily applicable software for the calculation of persistence indicators are provided. Chapters 3 and 4 address the representativeness of the laboratory-based OECD 308 and 309 simulation tests to reflect and predict the chemicals' fate in actual surface water bodies ("lab-to-field extrapolation"). For this purpose, Chapter 3 presents a literature review of the major factors influencing chemical fate in surface water bodies and how these are reflected in different laboratory-based test systems. It also provides a comparison of half-life data reported in the scientific literature for different laboratory-based systems.. Chapter 4 finally addresses lab-to-field extrapolation by comparing experimental data for actual in-field half-lives with half-lives derived from OECD 308 test outcomes. In the first section, again, half-lives in surface water systems from both field and laboratory studies as reported in the scientific literature are compared for a number of chemicals for which such data are available. In the second section, a specific case study is presented where monitoring data from the river Rhine was combined with experimental data from OECD 308 studies and a chemical fate model to assess the possibility to use OECD 308 data for exposure modelling. Finally, a summary of conclusions and recommendations is presented in Chapter 5.

# 2. Persistence indicators from OECD 308 & 309

## 2.1 Derivation of half-life indicators from OECD 308 and OECD 309 studies

Half-lives are the most important endpoints to be derived from OECD 308 and 309 experiments: they are both used for persistence classification and exposure modelling. Different half-lives are known and used in practice. Most directly observable is the total system half-life (abbreviate by different authors as $DissT_{50,ts}$ or $DegT_{50,ts}$), that is the disappearance half-life of the parent compound (P) from the entire experimental system. This can be computed right from the observed overall decline of P across all compartments with different kinetic models (FOCUS 2014). Despite being a disappearance half-life, $DegT_{50,ts}$ is also a meaningful degradation half-life due to the fact that the parent compound – except for highly volatile compounds – cannot escape the experimental systems unless transformed. In analogy to $DegT_{50,ts}$, disappearance half-lives can be defined for the water and sediment compartments in OECD 308 ($DissT_{50,w}$ and $DissT_{50,sed}$, respectively). However, these are known to merge biotransformation and phase transfer and are therefore not suitable to characterize the biotransformation properties of the subject compound. (Honti et al. 2015, ECHA 2016b). More meaningful compartment-specific half-lives are the true degradation half-lives in water and sediment ($DegT_{50,w}$ and $DegT_{50,sed}$, respectively), but these need to be derived indirectly from the observations by fitting a mechanistic model that describes both phase transfer and biotransformation pathways.

Based on these theoretical considerations of the various half-lives, we excluded compartment-specific disappearance half-lives ($DissT_{50,w}$ and $DissT_{50,sed}$) from our analysis due to the fact that phase transfer often dominates these values. The remaining $DegT_{50,ts}$, $DegT_{50,w}$ and $DegT_{50,sed}$ can all be considered better indicators of biotransformation, albeit with different dependency on other properties of the experimental systems (Honti et al. 2015).

In this part, we describe the methodology of deriving half-life indicators from OECD 308 and 309 studies. Besides the actual derivation process, we also describe the data collection method from the experimental dossiers. The reason for this is partly to introduce the standardised data format we used during the calculation of half-lives, partly to explain the methodology we used to fill typical data gaps in the experimental reports.

### 2.1.1 Methods

#### 2.1.1.1 Substance selection

**Selection procedure for pharmaceuticals**

*Selection criteria and method*

At the onset of the project, UBA had access to OECD 308 data for 112 pharmaceuticals. Within this project, the goal was to select 10-12 of these substances for data analysis. Highest priority was given to selecting 2-3 substances that were suitable for the case study presented in Chapter 4 (see Chapter 4.2.1.1 for details on their selection). 10 additional pharmaceuticals were then additionally selected according to the following criteria (in decreasing order of importance):

- Data quality
- Coverage of a wide spectrum of partitioning behaviour ($K_{OC}$) and $DegT_{50,ts}$
- Usage in Germany
- Presence of dominant, single transformation product (TP)

The selection method was not designed for a rigorous control of credibility of experimental data. As the target was to select 10 additional substances out of 112, we could not and did not plan to filter the datasets on data quality alone. While datasets having any trace of uncommon/special behaviour (e.g., dynamics that will not be described by the model, such as reactivation of NER, decrease in cumulated $CO_2$, fluctuations in transformation product amounts) or issues with data presentation were immediately excluded, this did not mean that their content was invalid or flawed. A solid judgment on the suitability of an experimental dataset for deriving persistence indicators can only be made after carrying out the estimation procedure and analysing the robustness and uncertainty of the resulting indicators. As it was out of the scope of this project to analyse all datasets quantitatively, we had to narrow down the number of datasets based on heuristic criteria.

112 pharmaceuticals were classified into priority classes, which we took as a starting point for selection. A first group of priority substances (priority class I) had been designated based on their consumption data or their supposed persistence (22 compounds in 24 experimental dossiers). Priority class II contained 22 substances with short dissipation half-life in water ($DissT_{50,w}$), but a long degradation half-life in the total system ($DegT_{50,ts}$), pointing towards substances with rather strong sorption and slow overall degradation. A single substance belonged to priority class III, which was characterized by the TP being more persistent than the parent compound. The majority of datasets (70) was not assigned any priority.

We used the priority classes as a basis for narrowing down the selection of pharmaceuticals to be studied. We bootstrapped the list of UBA priority substances by following an iterative selection algorithm:

1. We calculated $DissT_{50,w}/DegT_{50,ts}$ as a composite indicator of persistence and partitioning (Figure 3) from half-lives reported for the experimental studies.
2. We assessed the range of $DissT_{50,w}$, $DegT_{50,ts}$, and $DissT_{50,w}/DegT_{50,ts}$ values covered by priority class I compared to the range of all compounds having 308 data. $DissT_{50sed}$ was not included in the range calculations because it was not reported in the majority of studies.
3. We selected complementary substances from priority class II having valid dissipation half-lives to extend the set of selected compounds to about 40 and to broaden the range of $DissT_{50,w}$, $DegT_{50,ts}$, and $DissT_{50,w}/Deg_{T50,ts}$ where necessary.
4. We inspected the experimental datasets and classified them into 3 groups: A: good quality data; B: minor issues; C: serious issues. Data quality was mainly assigned based on two aspects. First, the mass balances of the experiments were checked. Recoveries below 90% and above 110% were generally considered problematic. Second, a qualitative assessment of the noisiness and the credibility of dynamics of individual chemical species was made based on the capabilities of the mathematical model of (Honti et al. 2015). Datasets with numerous missing components or problematic recoveries were classified as grade C. Strong noisiness (strong enough to make perceiving the dynamics of certain individual species difficult), multiple outliers and smooth yet strange dynamics resulted in grade B.
5. We deselected experimental datasets having a data quality grade C.
6. We reanalysed the remaining data according to point 2 to ensure that the selection still covers a reasonable proportion of compound properties.

Figure 1: Expected ranges for the R= $DissT_{50,w}$ / $DegT_{50,ts}$ ratio for different compound traits.

| $R = DissT_{50,w} / DegT_{50,ts}$ | Weakly sorbing | Strongly sorbing |
|---|---|---|
| Non transforming | $DissT_{50,w}$ long<br>$DegT_{50,ts}$ long<br>$R \approx I$ | $DissT_{50,w}$ short<br>$DegT_{50,ts}$ long<br>$R \ll I$ |
| Transforming | $DissT_{50,w}$ short<br>$DegT_{50,ts}$ short<br>$R \leqq I$ | $DissT_{50,w}$ short<br>$DegT_{50,ts}$ short<br>$R < I$ |

*Selection results*

Substances belonging to Priority class I already covered most of the range of $DissT_{50,w}$, $DegT_{50,ts}$, and $DissT_{50,w}/DegT_{50,ts}$ values of all OECD 308 datasets (Table 1). Therefore, all 24 datasets belonging to priority class I were included in the first round of selection. As class I datasets already provided a sufficient coverage of parameter ranges, this was not a major criterion for inclusion of class II data sets anymore. Instead, those 17 substances out of 22 class II substances were added that contained a valid and definite $DegT_{50,ts}$ value. This yielded altogether 41 experimental studies to be analysed for data quality.

Table 1: Parameter ranges covered by the entire dataset and the experiments belonging to priority class I.

| Parameter | Range in entire dataset | Range in priority class I |
|---|---|---|
| $DissT_{50,w}/DegT_{50,ts}$ | 0.002-1.846 | 0.002-1.353 |
| $DissT_{50,w}$ [d] | 0.15-176 | 0.99-176 |
| $DegT_{50,ts}$ [d] | 0.46-2310 | 6.59-2310 |
| $DissT_{50,w}/DegT_{50,ts}$ | 0.002-1.846 | 0.002-1.353 |
| $DissT_{50,w}$ [d] | 0.15-176 | 0.99-176 |

Datasets with C data quality grade (unreported important data, bad recoveries, strange degradation dynamics, etc.) were excluded from the selection, and 33 datasets (20 from priority class I, 13 from priority class II) remained. To narrow down the set of pharmaceuticals even further to ten substances, besides those compounds required by WPII, we excluded datasets with quality grade B (better data quality than C grade, but also problematic at some points). We further focused on priority class I compounds because they mostly covered a similar range of $DissT_{50,w}/DegT_{50,ts}$ values as the substances in priority class II, and were additionally important because of high usage and/or persistence. The final selection contained 14 datasets for 13 compounds (Table 2).

Table 2: Pharmaceutical datasets included in final selection.

| # | Compound | Priority class | Data quality grade |
|---|---|---|---|
| 1 | API1 | I | A |
| 2 | API2 | I | A |
| 3 | API3 | I | A |
| 4 | API4 | I | A |
| 5 | API5 | I | A |
| 6 | API6 | I | B |
| 7 | API7 | I | A |
| 8 | API8 | I | A |
| 9 | API9[1] | I | A/B |
| 10 | API9[2] | I | A |
| 11 | API10 | I | A/B |
| 12 | API11 | I | A |
| 13 | API12 | I | A |
| 14 | API13 | I | B |

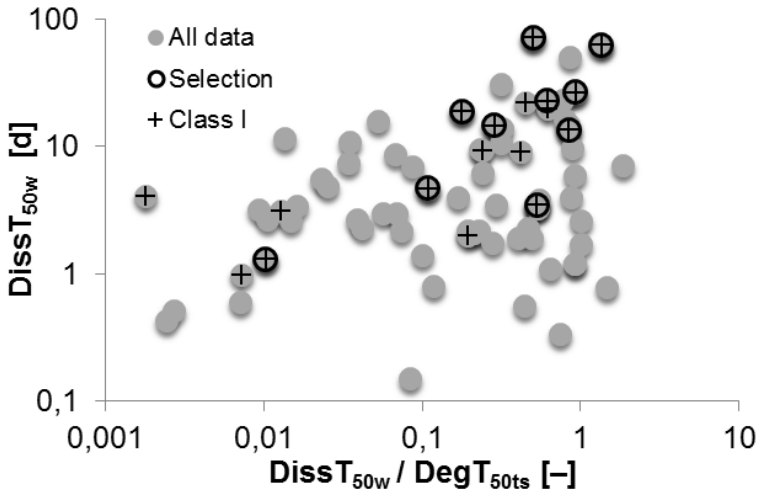
[1,2] There are 2 reported studies selected for API9.

The final selection still covered a reasonable range in terms of $DissT_{50,w}/DegT_{50,ts}$, $DegT_{50,ts}$ and $DissT_{50,w}$ (Table 3, Figure 2). The range covered in $DissT_{50,w}$ has been reduced from 176 to 72 days, but this was only because the only compound with a $DissT_{50,w}$ above 80 days in the entire dataset was not included anymore due to data quality problems. The maximal featured value for $DegT_{50,ts}$ has reduced to 287 days, but this is still well beyond the usual duration of OECD 308 experiments and the range covered can therefore still be considered a proper representation of persistence values that can be confidently determined in 308 studies.

Table 3: Parameter ranges (from reports) covered by the entire dataset versus the selected pharmaceuticals.

| Parameter | Range in entire dataset | Range in selection |
|---|---|---|
| $DissT_{50,w}/DegT_{50,ts}$ | 0.002-1.846 | 0.010-1.353 |
| $DissT_{50,w}$ [d] | 0.15-176 | 1.3-72 |
| $DegT_{50,ts}$ [d] | 0.46-2310 | 6.59-287 |

Figure 2: Distribution of all data, priority class I compounds and the final selection in the $DissT_{50,w}$ vs. $DissT_{50,w}$ / $DegT_{50,ts}$ space. Four selected class I compounds (API3, API4, API5, and API12) did not have valid $DissT_{50,w}$ or $DegT_{50,ts}$ values reported and hence are not shown.




**Selection procedure for pesticides**

It was decided to prioritize those pesticides for analysis for which both OECD 308 and OECD 309 data are available. This is the case for 14 pesticides, which are listed in Table 4 along with the data quality scores for both OECD 308 and 309 data.

Table 4: Pesticides with both OECD 308 and OECD 309 data.

| # | Compound | Water-sediment study data quality (kind) | OECD 309 data quality (kind) |
|---|---|---|---|
| 1 | PE1 | B/C (OECD 308) | A (pelagic) |
| 2 | PE2 | B (EFS092, EFS103) | A (pelagic) |
| 3 | PE3 | A (OECD 308) | A (pelagic) |
| 4 | PE4 | A (OECD 308) | A (pelagic) |
| 5 | PE5 | B (OECD 308) | A (pelagic) |
| 6 | PE6 | B (OECD 308) | A (pelagic) |
| 7 | PE7 | A (OECD 308) | A (pelagic) |
| 8 | PE8 | A (OECD 308) | A (pelagic) |
| 9 | PE9 | A (OECD 308) | A (pelagic) |
| 10 | PE10 | A (OECD 308) | A (pelagic) |
| 11 | PE11 | B (95/36/EC) | B (pelagic/sus-pended; light/dark) |
| 12 | PE12 | A (German BBA,Part IV,5-1) | B (suspended) |

| # | Compound | Water-sediment study data quality (kind) | OECD 309 data quality (kind) |
|---|---|---|---|
| 13 | PE13 | A (German BBA,Part IV,5-1) | A (pelagic) |
| 14 | PE14 | A (OECD 308) | B (pelagic/suspended) |

#### 2.1.1.2 Data collection from experimental dossiers

Necessary data for half-life calculations were gathered in a standardised data format from all experiments, regardless of experimental type. The two elements of this data format were a residue table describing the time-dynamics of compartment-specific radioactivity of the parent compound, all transformation products pooled together, $CO_2$, non-extractable residues, and an additional list of experimental metadata.

**Residue tables**

Residue tables all had the following columns:

- Time: sampling days [d]
- $P_w$: Radioactivity in water compartment from the parent compound [% of applied radioactivity]
- $M_w$: Radioactivity in water compartment from the sum of all transformation products [% of applied radioactivity]
- $P_s$: Radioactivity in sediment compartment from the parent compound [% of applied radioactivity]
- $M_s$: Radioactivity in sediment compartment from the sum of all transformation products [% of applied radioactivity]
- $CO_2$: $CO_2$ caught in traps [% of applied radioactivity]
- NER: non-extractable or bound residue [% of applied radioactivity]

For missing or not applicable data we used 'NA' as a placeholder. When a component was not detectable or quantifiable during a measurement, zero value was used in the residue table, because detection and quantification limits were typically very low (<0.1%) compared to the applied radioactivity (Table 5).

Table 5: Example for residue table (compound: API8). $P_w$: Radioactivity in water phase from the parent compound, $M_w$: Radioactivity in water compartment from all transformation products, $P_s$: Radioactivity in sediment compartment from the parent compound, $M_s$: Radioactivity in sediment phase from all transformation products, $CO_2$: $CO_2$ caught in traps, NER: non-extractable or bound residue [% of applied radioactivity].

| Time | $P_w$ | $P_s$ | $M_w$ | $M_s$ | $CO_2$ | NER |
|---|---|---|---|---|---|---|
| 0 | 99.20 | NA | 0.00 | NA | NA | 0.0 |
| 2 | 81.30 | 15.90 | 2.90 | 0.00 | 0.00 | 0.2 |
| 7 | 73.60 | 24.50 | 2.40 | 0.30 | 0.00 | 0.6 |
| 14 | 64.80 | 26.60 | 3.40 | 0.50 | 0.00 | 1.3 |
| 29 | 58.80 | 31.00 | 5.80 | 1.30 | 0.50 | 1.5 |

| Time | $P_w$ | $P_s$ | $M_w$ | $M_s$ | $CO_2$ | NER |
|---|---|---|---|---|---|---|
| 58 | 45.40 | 34.00 | 11.80 | 4.80 | 0.50 | 6.8 |
| 98 | 36.20 | 25.80 | 16.00 | 7.00 | 0.50 | 11.1 |
| 230 | 16.10 | 15.00 | 23.60 | 9.70 | 5.00 | 28.6 |

To reduce data scatter, the mean values of the data from experimental duplicates were used to produce the residue tables.

For the joint modelling of different experiments belonging to the same compound and same sediment (section 2.1.1.6), joint residue tables had to be assembled from the individual experiments' residue tables. To facilitate this process, we harmonized the measurement days already during the compilation of the individual residue tables: all tables belonging to the same compound and sediment featured the same time coordinates, with full 'NA' rows used for data not measured in a given experiment.

**Experimental Metadata**

Modeling OECD 308 or 309 systems is difficult based on a residue table alone. Both experimental systems, although 309 to a lesser degree, are so complex that processes are difficult to identify just based on the observed time-dynamics of the parent compound and transformation products. This is actually the very same reason why compartment-specific dissipation half-lives are poor indicators of biotransformation in OECD 308: different combinations of the extent of phase transfer and biotransformation lead to the same dissipation pattern, and therefore neither the rate nor the compartment where biotransformation takes place can be identified from $DissT_{50,w}$ and $DissT_{50,sed}$. The problems of system complexity certainly affect $DegT_{50,w}$ and $DegT_{50,sed}$ too, as they are even more indirectly related to the actual observed data. This identification problem can be tackled by utilising additional (meta)data besides the residue table (Honti et al. 2015). These metadata describe the experimental boundary conditions and other relevant environmental fate properties of the compound that need to be known from independent experiments or prediction.

Collected metadata belonged to two main categories. The first category contained qualitative information necessary for the identification of experiments and correct interpretation of results. The metadata in this category were the following:

- Name of the compound
- Sampling site of water (and sediment) used in the experiment
- Redox conditions during the test: aerobic or anaerobic test?
- Applied dose of the compound (default, high, or low)
- Test system (OECD 308 or OECD 309)

The second class of metadata contained quantitative information that was effectively used during the calculation of half-lives. Quantitative metadata were the following for all experimental types:

- Sediment organic carbon-water partition coefficient ($K_{oc}$ [L $kg^{-1}$])
- Relative $K_d$ of transformation products ($dK_d=K_{d,M} / K_{d,P}$ [–])
- Total organic carbon concentration in water ($TOC_w$ [mg $L^{-1}$])
- Dissolved organic carbon concentration in water ($DOC_w$ [mg $L^{-1}$])
- Is the compound hydrolysing? (yes or no)

For OECD 308 experiments and non-pelagic OECD 309 experiments, the organic carbon content of the applied sediment was important:

- Sediment organic carbon content ($f_{oc,sed}$ [%])

System geometry (compartment heights after incubation) was crucial for OECD 308 experiments, along with the porosity of the wet sediment, which was calculated from the geometry:

- Height of wet sediment, preferably after incubation ($Z_s$ [cm])
- Height of water column, preferably after incubation ($Z_{wc}$ [cm])
- Porosity of wet sediment, preferably after incubation ($\theta$ [–])

The suspended solids concentration was collected for non-pelagic OECD 309 experiments:

- Total suspended solids concentration (TSS [kg $L^{-1}$])

Parameter priors for Bayesian parameter inference and uncertainty analysis were defined based on the available metadata. We tried to gather mean and standard deviation values for each metadata variable. Means were set equal to exact measured values where such data were available. Standard deviations were assumed to be relative to the measured mean in such cases. In case of multiple available measured values, without any of them being clearly superior, population means and standard deviations were calculated (e.g. $K_{oc}$ for different soils/sediments). Missing metadata were set to estimated values (details in the next section) with relatively high associated uncertainty.

#### 2.1.1.3 Handling missing and erroneous data

Data collection faced numerous difficulties due to the variety of report formats and kinds of information presented. Many reports followed a structure that mirrored measurement stages exactly, which made it difficult to assemble a residue table with closed mass balance.

The problem in these cases was that columns of the residue table come from different measurement stages and therefore the complete value set of a given row is not to be found in any published table and the published mass balances do not apply to the residue table.

While missing data were of less concern for residue tables, certain required metadata were almost never reported. These included system geometry, the porosity of wet sediment, and $K_{oc}$ for the exact sediment used in the studies. To overcome the problems of missing metadata, we relied on expert estimates partly based on the OECD 308 standard (for example: system geometry) and partly on common knowledge (for example: porosity, $dK_d$) for the non-reported properties. The assumptions used to quality control reported data and to fill in missing data will be described below.

**Metadata estimation policy**

*Sediment-water partition coefficient ($K_d$):*

In the OECD 308 and 309 reports, $K_d$ values are typically not given for the specific sediment used in the experiments. To come up with a statistical estimate for the experimental $K_d$, we collected $K_{oc}$ values from OECD 106, FDA 3.08, OPPTS 835.1110, and OECD 121 studies for each compound. If the partition coefficients were available for soils, we used those, if not, we relied on values for sludges. The mean and standard deviation of $K_d$ were calculated from these $K_{oc}$ and the experimental $f_{oc,sed}$ values. If there was only one $K_{oc}$ value available, we assumed $K_{oc}$* $f_{oc,sed}$ to be the mean sediment-water partition coefficient, and the standard deviation was set to 114% of the mean. If there were only two values, the standard deviation was assumed to be their standard deviation, but at least 85% of their mean value. For 3 or more values the standard deviation was estimated via the normal procedure.

*Relative sediment-water partition coefficient of transformation products ($dK_d$):*

This property is rather abstract, as we defined $dK_d$ as the statistically representative relative $K_d$ of the lumped transformation product pool. A proper estimation of this would require knowing the exact $K_d$ of the parent compound for the given sediment, the exact $K_d$ of all transformation products produced during the experiment and even their relative quantities. This predestined this value to be never reported for pharmaceuticals

and only for some of the PPPs. Moreover, the reasonably precise estimation of $dK_d$ from other reported data would require specialised chemical knowledge. Therefore, we assumed that while the relative $K_d$ of transformation products is very uncertain, they generally tend to be somewhat more polar compared to the parent compound and hence used a mean value of 0.8 along with a standard deviation of 0.4.

*Sediment organic carbon content ($f_{oc,sed}$):*

Preferably, we used the sediment organic carbon content from the start of the incubation time, wherever it was reported. If it was not specified, we took a value from during the incubation time or from the acclimation period. If there was no data from the laboratory systems, we used the sediment organic carbon content measured at the sampling site. In some cases, only organic matter content ($f_{om,sed}$) was reported, so we had to convert it to $f_{oc,sed}$ by assuming a certain stoichiometry (OC [%] = OM [%] / 1.7).

*Wet sediment depth ($Z_{sed}$):*

If the depth of wet sediment in the flask was not specified in the report, we assumed it to be 3 cm, and the standard deviation to be 1 cm to cover all allowed valuess from the OECD 308 guideline. For reported wet sediment depth values we took 0.1 cm as standard deviation.

*Water column height ($Z_{wc}$):*

If the water column height in the flask was not specified in the report, we assumed it to be 6 cm, and the standard deviation to be 2 cm, again to cover all allowed values from the OECD 308 guideline. For reported water column height values we took 0.1 cm as standard deviation.

*Porosity of the wet sediment ($\theta$):*

Porosity of the wet sediment in OECD 308 is a key determinant of the sorption and diffusion behaviour inside the settled sediment layer, but it was generally not reported. Therefore, we tried to derive estimates from other system properties that were typically better reported. Two base equations were used for porosity calculation (equations 1 and 2) based on data availability:

$$\theta = 1 - (M_{sed,dry}\ [g] \,/\, V_{sed,wet}\ [cm^3])/2.5\ [g\ cm^{-3}] \qquad (1)$$

where $M_{sed,dry}$ is the mass of the dry sediment , 2.5 [g $cm^{-3}$] is the estimated density of sediment solids, and the wet volume of sediment ($V_{sed,wet}$) is calculated from:

- $Z_s$ [cm] × area of the flask's bottom [$cm^2$] or
- from reported wet sediment:water volume ratio

In case these data were not available, we used

$$\theta = V_{porewater}\ [cm^3] \,/\, V_{sed,wet}\ [cm^3] \qquad (2)$$

where the volume of porewater in the settled sediment layer ($V_{porewater}$) is calculated from:

- (Wet sediment weight – Dry sediment weight [$cm^3$]) / 1 [g $cm^{-3}$]

In many cases, porosity calculation was impossible, as lack of data prohibited the use of both potential equations. In these cases, we assumed the porosity to be 0.7 with 0.15 standard deviation. If there was enough data for porosity calculation, we applied 0.1 for standard deviation.

*Suspended solids concentration (TSS):*

In non-pelagic OECD 309 tests, we used the reported concentrations of suspended solids [kg $L^{-1}$], and we assumed the standard deviation to be 20% of the reported value.

*Total ($TOC_w$) and dissolved organic carbon ($DOC_w$) concentration of water column:*

We tried to use TOC and DOC concentrations measured at the start of the incubation period. If it was not given, we took values from during the experiment, or from the acclimation period. If there was no data from the laboratory, we used the TOC and DOC values measured at the sampling site. In some cases, either TOC or DOC, or both of them were missing from the reports, so we had to use estimations. If only one of them was available, we applied the following assumptions for the missing value:

- $TOC_w$ = 140% of $DOC_w$ when not reported.
- $DOC_w$ = 60% of $TOC_w$ when not reported.

If neither TOC nor DOC value were reported, we assumed them to be 7 and 4 mg $L^{-1}$ , respectively. Standard deviations were assumed to be 30% of the mean for reported values and 70% of the mean for non-reported values.

**Residue table quality control policy**

Measurement records, that is values for a given measurement day, with a recovery above 120% or below 80% were excluded from analysis, unless the specific measurement record was a crucial point determining the general shape of the time series. Examples of crucial points include starting, central, or final time points without a close neighbouring record of good quality.

#### 2.1.1.4 Calculation of total system half-life ($DegT_{50,ts}$)

Total system half-life was calculated by fitting the semi-empirical kinetic models recommended by the FOCUS Kinetics Report (FOCUS 2014) to the overall amount of the parent compound in the entire experimental system. These models were re-implemented by us so that the same calibration and uncertainty assessment procedure used for compartment-specific half-lives can also be carried for $DegT_{50,ts}$.

The simplest model is the 'Single First-Order' (SFO) model with the initial amount ($M_0$) and the first-order degradation rate constant ($k_{SFO}$) as parameters:

$$M(t) = M_0 \exp(-k_{SFO}\, t) \qquad (3)$$

where M(t) is the amount (or radioactivity) of the parent compound at time t. $DegT_{50,ts}$ derives analytically from the previous equation by setting $M(t) = 0.5\, M_0$:

$$DegT_{50,ts} = \frac{\ln 2}{k_{SFO}} \qquad (4)$$

A more complex model is the 'Double First-Order in Parallel' (DFOP) model. Here, two first-order degradation processes are happening simultanously on different parts of the initial mass, determined by a partitioning factor $g_{DFOP}$:

$$M(t) = g_{DFOP}\, M_0 \exp(-k_{1DFOP}\, t) + (1 - g_{DFOP}) M_0 \exp(-k_{2DFOP}\, t) \qquad (5)$$

where $k_{1DFOP}$ and $k_{2DFOP}$ are the degradation rates of the two simultaneous processes [$d^{-1}$]. Despite the simple structure of this model, the actual $DegT_{50,ts}$ cannot be derived analytically. For this model, we followed the usual practice of reporting the half-life of the slower degradation process according to equation (4).

A similarly complex alternative was the 'Hockey-Stick' (HS) kinetical model, where partitioning between two distinct processes is not in terms of mass but in terms of time:

$$M(t) = \begin{cases} M_0 \exp(-k_{1HS} \cdot t) & \text{if } t < t_{HS} \\ M_0 \exp(-k_{1HS} \cdot t_{HS}) \exp\left(-k_{2HS}(t - t_{HS})\right) & \text{if } t > t_{HS} \end{cases} \qquad (6)$$

where $k_{1HS}$ and $k_{2HS}$ are the degradation rate constants of the two sequential processes [$d^{-1}$], and $t_{HS}$ is the time divider between the two processes [d]. $DegT_{50,ts}$ of the HS model is calculated based on the ability of the first process to consume more than the half of the parent compound:

$$DegT_{50,ts} = \begin{cases} \frac{\ln 2}{k_{1HS}} & \text{if } k_{1HS} \cdot t_{HS} > \ln 2 \\ -\frac{\ln(0.5 - \exp(-k_{1HS} \cdot t_{HS}))}{k_{2HS}} & \text{if } k_{1HS} \cdot t_{HS} < \ln 2 \end{cases} \quad (7)$$

The 'First-Order Multi-Compartment' (FOMC) model of the FOCUS report was not used due to mathematical stability reasons.

#### 2.1.1.5 Calculation of compartment-specific biotransformation half-lives ($DegT_{50,w}$ and $DegT_{50,sed}$)

The calculation of compartment-specific biotransformation half-lives requires a mechanistic model of the experimental system, as biotransformation in water and sediment is not directly observable. A model framework developed by (Honti et al. 2015) was used for this purpose. This model framework is capable of simulating both OECD 308 and OECD 309 experiments and therefore allows cross-experiment calibration, that is the estimation of biotransformation rates from multiple experiments at once. The following description of the model framework is based on the supporting information for (Honti et al. 2016).

The basic building block of the model framework is a fully-mixed reactor layer containing both water and sediment. The following assumptions are made:

1. Sorption takes place to particles, including colloids.
2. The dynamics of sorption can be described by a first-order convergence towards equilibrium.
3. Particulate organic matter is a good proxy for the available biomass, assuming that the fraction of active degraders within the particulate organic matter is the same in all three reactor types.
4. Transformation products are merged into a unified pool.

Based on these assumptions, one can formulate the mathematical model for such a reactor. Following assumption 1, the equilibrium concentrations are calculated as follows for both parent compound (P) and pooled transformation products (M) (shown for parent):

$$\hat{P}_{w,eq} = \hat{P}_{total} \frac{1}{1 + K_d \, TSS \frac{V_{tot}}{V_{aq}}} \quad (8)$$

where $\hat{P}_{w,eq}$ is the equilibrium concentration of the parent compound in water [kg $L^{-1}$], $\hat{P}_{total}$ is the total amount of the parent compound in the system [kg $L^{-1}$], $K_d$ is the sediment solids-water partition coefficient [L $kg^{-1}$], TSS is the total particle concentration [kg $L^{-1}$], $V_{tot}$ is the total reactor volume [L], and $V_{aq}$ is the aqueous volume [L].

According to assumption 2, the actual concentrations of the aqueous and sorbed fractions would follow a first-order kinetics in the absence of other (bio)chemical processes:

$$\frac{d\hat{P}_w}{dt} = -k_{sorp} \left( \hat{P}_{w,eq} - \hat{P}_w \right) \quad (9)$$

where $\hat{P}_w$ is the actual concentration of the parent compound in water [kg $L^{-1}$], and $k_{sorp}$ is the general sorption rate applying to both parent and transformation products [$d^{-1}$].

In line with assumption 3, biotransformation rate constants are considered to be a product of a second-order biomass-specific transformation rate constant $k'_{bio}$ [(kg OC $L^{-1})^{-1}$ $d^{-1}$]) and particulate organic carbon (POC = TOC – DOC) as a proxy for degrader biomass. Only the aqueous fractions ($\hat{P}_w$ and $\hat{M}_w$) are considered to be bioavailable, the particle-bound fractions ($\hat{P}_s$ and $\hat{M}_s$) are not available for biotransformation.

The general mass balance equations applying to all reactor types are the following:

$$\frac{d\hat{P}_w}{dt} = +k_{sorp} \left( \hat{P}_{w,eq} - \hat{P}_w \right) - k'_{bio,P} \, POC \cdot \hat{P}_w - k_{hydr} \hat{P}_w \quad (10)$$

$$\frac{d\hat{P}_s}{dt} = -k_{sorp}\left(\hat{P}_{w,eq} - \hat{P}_w\right) - k_{pn}\,\hat{P}_s \quad (11)$$

$$\frac{d\hat{M}_w}{dt} = +k_{sorp}\left(\hat{M}_{w,eq} - \hat{M}_w\right) + k'_{bio,P}\,POC\cdot\,\hat{P}_w + k_{hydr}\,\hat{P}_w - k'_{bio,M}\,POC\cdot\,\hat{M}_w \quad (12)$$

$$\frac{d\hat{M}_s}{dt} = -k_{sorp}\left(\hat{M}_{w,eq} - \hat{M}_w\right) - k_{mn}\,\hat{M}_s \quad (13)$$

$$\frac{dCO_2}{dt} = +k'_{bio,M}\,POC\cdot\,\hat{M}_w \quad (14)$$

$$\frac{dNER}{dt} = +k_{pn}\,\hat{P}_s + k_{mn}\hat{M}_s \quad (15)$$

where $k_{hydr}$ [$d^{-1}$] is the rate of hydrolysis (if applicable), $k_{pn}$ and $k_{mn}$ [$d^{-1}$] are the NER formation rates from parent and transformation products, respectively.

***Type I reactors: pelagic layer***

Type I reactors have low TSS and POC concentrations, and $V_{tot}/V_{aq}$ is very close to 1. POC can be calculated as the product of $f_{oc,sed}$ and TSS or by subtracting DOC from TOC. Equation (9) applies for sorption dynamics. Equations (10)-(15) can be used to describe the dynamics of state variables.

***Type II reactors: water with suspended sediment***

Type II reactors have medium TSS and TOC concentrations with both DOC and POC contributing to TOC. As sediment is scarce enough to be suspended by light stirring or shaking, $V_{tot}/V_{aq}$ is very close to 1. POC can be calculated as the product of $f_{oc,sed}$ and TSS. Equation (9) applies for sorption dynamics. Equations (10)-(15) can be used to describe the dynamics of state variables.

***Type III reactors: settled sediment with porewater***

The TOC content of type III reactors is deduced from sediment properties.

The bulk density of sediment ($\rho_b$ [kg $L^{-1}$]) can be calculated from porosity ($\theta$ [–]) and an assumed solid particle density of 2.5 [kg $L^{-1}$]:

$$\rho_b = 2.5\,(1-\theta) \quad (16)$$

and then

$$TOC_{III} = \rho_b \cdot f_{oc,sed} \quad (17)$$

In type III reactors we assume an instantaneous sorption equilibrium ($k_{sorp}=\infty$), so the aqueous fraction can be directly calculated from equation (8) at any time point:

$$\hat{P}_w = \hat{P}_{total}\,\frac{1}{1+K_d\cdot\frac{\rho_b}{\theta}} \quad (18)$$

and

$$\hat{P}_s = \hat{P}_{total} - \hat{P}_w \quad (19)$$

where $\hat{P}_{total}$ is the total amount of the parent compound in the reactor.

Due to the instantaneous sorption equilibrium, differential mass balance equations can only be formulated for the entire reactor layer:

$$\frac{d\hat{P}_{total}}{dt} = -k'_{bio,P}\,POC\cdot\,\hat{P}_w - k_{pn}\hat{P}_s - k_{hydr}\,\hat{P}_w \quad (20)$$

$$\frac{d\hat{M}_{total}}{dt} = +k'_{bio,P}\,POC\cdot\,\hat{P}_w + k_{hydr}\,\hat{P}_w - k'_{bio,M}\,POC\cdot\,\hat{M}_w - k_{mn}\hat{M}_s \quad (21)$$

$$\frac{dCO_2}{dt} = +k'_{bio,M}\,POC\cdot\,\hat{M}_w \quad (22)$$

$$\frac{dNER}{dt} = +k_{pn}\,\hat{P}_s + k_{mn}\hat{M}_s \quad (23)$$

As anaerobic conditions are common in deeper parts of settled sediments, equations (20)-(23) are slightly modified when this is the case. In anaerobic type III reactors, all biotransformation rates ($k'_{bio,P}$, $k'_{bio,M}$, $k_{pn}$, and $k_{mn}$) were considered to be uniformly lower than the aerobic ones. Transformation rates were multiplied by a relative anaerobic coefficient $dk_{aer}$ [–] in anaerobic reactors, with $0 < dk_{aer} < 1$.

**Description of the OECD 308 and OECD 309 systems with reactors**

Models for OECD 308 and 309 systems can be built by assembling reactors of type I, II, and III. OECD 309 systems are the simplest case due to their fully mixed nature. Pelagic or non-pelagic OECD 309 can be modelled by taking a single reactor of type I or type II, respectively. Then the reactor's amount of parent compound ($\hat{P}$) and transformation products ($\hat{M}$) are identical to the system's corresponding amounts (P and M). The first-order biotransformation rate in the water phase ($k_{wpm}$ [$d^{-1}$]) derives simply from equation (10):

$$k_{wpm} = k'_{bio,P}POC = k'_{bio,P}\left(f_{oc,sed}TSS + TOC_w - DOC_w\right) \quad (24)$$

and from there $DegT_{50,w}$ by assuming first-order overall decay becomes:

$$DegT_{50,w} = \frac{\ln 2}{k_{wpm}} \quad (25)$$

OECD 308 experiments are more complex. The sediment compartment is predominantly anaerobic (Honti et al. 2015, Shrestha et al. 2016) with a thin aerobic surface layer where the majority of biotransformation usually takes place. This means that – besides a type I reactor for the non-stirred and therefore pelagic water column – multiple reactors are necessary for a realistic representation of sediment. In line with the model definition in Honti and Fenner (2015) and Honti et al. (2016), we used 4 stacked sediment reactors of downwards increasing depth (6.7%, 13.3%, 26.7%, and 53.3% of the total sediment depth), with only the thinnest and upmost one being aerobic. These stacked reactor layers (1 for water phase, 1 aerobic sediment, 3 anaerobic sediments) are linked by diffusion processes between adjacent elements. Having multiple sediment layers of downward increasing depth provides a realistic and robust numerical simulation of the exponential concentration gradients that typically build up in settled sediment layers.

During the calculation of diffusion between the type I layer and the uppermost type III layer we neglect the depth of the type I layer ($Z_I$=$Z_{wc}$) from the diffusion distance because of the much faster diffusion in a fully mixed water body compared to tortuous diffusion in a settled sediment layer. The diffusive flux is calculated from the concentration gradient and the diffusion distance as follows:

$$\Psi_{D_{I\,to\,III}} = A\cdot D_P\frac{C_I - C_{III}}{d} = A\cdot D_P\frac{\left[\frac{\hat{P}_{w_I}}{Z_I}\right] - \left[\frac{\hat{P}_{w_{III}}}{Z_{III}}\right]}{\frac{Z_{III}}{2}} \quad (26)$$

where $\Psi_{D_{I\ to\ III}}$ is the diffusive flux [radioactivity $d^{-1}$] from the type I layer to the type III layer, A is the surface area of the layer [$cm^2$], $D_P$ is the diffusion coefficient (now for the parent compound) [$cm^2\ d^{-1}$], $C_I$ and $C_{III}$ are the aqueous concentrations in the two reactors, d is the diffusion distance, and $Z_I$ and $Z_{III}$ denote the depths of the two layers [cm]. For diffusion between two type III layers (identified by indices a and b, with depths of $Z_{IIIa}$ and $Z_{IIIb}$, respectively), we explicitly consider the distance between the centerlines of layers:

$$\Psi_{D_{IIIa\ to\ IIIb}} = A \cdot D_P \frac{\left[\frac{\hat{P}_{w_{IIIa}}}{Z_{IIIa}}\right] - \left[\frac{\hat{P}_{w_{IIIb}}}{Z_{IIIb}}\right]}{\frac{Z_{IIIa}+Z_{IIIb}}{2}} \quad (27)$$

These fluxes are added to the mass balance equations of $\hat{P}_w$ and $\hat{M}_w$ in type I layers and $\hat{P}_{total}$ and $\hat{M}_{total}$ in type III layers when used to simulate an OECD 308 system. Contrary to the OECD 309 experiment, here the aqueous and particle-bound $\hat{P}$ and $\hat{M}$ fractions of the single layers do not map directly to system-level concentrations. Instead, $\hat{P}_{total}$ of the (type I) water layer is the equivalent of the radioactivity in water, and $\hat{P}_{total}$ of all (type III) sediment layers is the equivalent of the radioactivity in sediment. $\hat{P}_w$ of the type III layers is actually the seldom measured radioactivity in porewater and $\hat{P}_s$ is the truly particle-bound fraction.

The first-order biotransformation rate in the water phase in OECD 308 is then:

$$k_{wpm} = k_{bio,P} \frac{TOC_w - DOC_w}{1 + K_d \cdot TSS} \quad (28)$$

and the first-order biotransformation rate in the topmost aerobic layer of sediment is:

$$k_{spm} = k_{bio,P} \frac{f_{oc,sed}}{1 + K_d \cdot \frac{\rho_b}{\theta}} \rho_b \quad (29)$$

In the deeper, anaerobic sediment layers the corresponding first-order biotransformation rate is $k_{spm} \cdot dk_{aer}$.

The compartment-specific biotransformation half-lives are calculated from the first-order rates:

$$DegT_{50,w} = \frac{\ln 2}{k_{wpm}} \text{ and } DegT_{50,sed} = \frac{\ln 2}{k_{spm}} \quad (30)$$

#### 2.1.1.6 Calibration and uncertainty analysis of models

All indicators ($DegT_{50,ts}$, $DegT_{50,w}$, $DegT_{50,sed}$) derive from fitting a model to the observed experimental data. The fitted model is rather simple for $DegT_{50,ts}$ with the SFO assumption and quite complicated for $DegT_{50,w}$ and $DegT_{50,sed}$ in case of an OECD 308 experiment, but the procedure is essentially the same: Model parameters need to be adjusted so that the model's simulations match observations as closely as possible, then the desired half-lives can be derived from the model parameters. A parameter set that leads to the smallest possible discrepancies between the observed and simulated quantities can be considered as optimal. Due to the errors in observations and in model structure, the optimal parameter set will still have non-zero discrepancies between observed and simulated values, so optimality does not mean perfect – undoubtable – fit. While for each calibration problem there exists only one optimal error level, it can belong to different model outputs. As the best error level is non-zero, multiple model simulations can have the same error level (for example, one solution fits better a certain part of the observations, worse in the rest, while another solution does it vice versa). For more complex models it even happens that two or more distinct and dissimilar parameter sets lead to exactly the same model output and error. This possibility for multiple parameter sets being optimal justifies the need for uncertainty assessment. In an uncertainty assessment, a high variety of parameter combinations is evaluated, and besides the optimal values the statistically not significantly worse solutions are saved too. In contrast to traditional model calibration, the result of the uncertainty assessment is not a single set

of model parameters leading to the lowest possible error but rather a set of parameter distributions that statistically describe the neighbourhood of the optimal values.

For $DegT_{50,ts}$ the goodness of model fit was calculated with a formal statistical likelihood based on the classical assumption of independent, and normally-distributed errors:

$$\mathcal{L} = \prod_{i=1}^{n} \frac{1}{\sqrt[2]{2\pi\sigma}} \exp\left(-\frac{(Y_{m,i}-Y_{o,i})^2}{\sigma^2}\right) \quad (31)$$

where $\mathcal{L}$ is the likelihood of a certain parameter combination and the corresponding model fit, n is the number of observation points, σ is the standard deviation of the independent, normally distributed model error, $Y_{m,i}$ and $Y_{o,i}$ are the modelled and observed values (% aR) for the $i^{th}$ observation point. This, if one considers the log-likelihood as an error indicator, leads to the well-known least squares optimization.

For the compartment-specific half-lives the complexity of the model to be fitted required the use of additional information besides the residue table. We therefore used Bayesian calibration and uncertainty assessment. The goodness of fit indicator in this case is called the posterior probability, which is the product of the above-mentioned likelihood from equation (32) and the prior probability of parameter values:

$$P \propto p\,\mathcal{L} \quad (32)$$

where P is the posterior probability of a parameter combination and the corresponding model fit, ∝ indicates proportionality, and p is the prior probability of the parameter combination calculated from the prior distributions for each parameter by multiplying the individual prior probabilities. The involvement of the prior probabilities, which do not depend on the actual observations (i.e. the residue table), means that the objective of Bayesian calibration will not be solely to attain the best possible fit to the residue table. Instead, a compromise solution will be sought where the fit to observations and conformity to the expectations expressed by the prior distributions are fulfilled simultaneously as much as possible. Such a construct does certainly compromise the model fit, that is the model error will be higher or at best the same compared to calibration based on likelihood only. However, in case of weakly identifiable models such as the mechanistic model used to calculate $DegT_{50,w}$ and $DegT_{50,sed}$, the price of this compromise is very low, i.e., the maximum posterior probability solution is not significantly worse from the maximum likelihood solution in terms of model fit. Yet the Bayesian approach helps to restrict the internal, normally non-identifiable model processes to realistic domains and therefore to get meaningful and less uncertain parameters.

Models were fitted to the averaged data of the two duplicates for each experiment. Parameter priors were defined according to the procedures described in section 2.1.1.3. Calibration was performed with Nelder-Mead Simplex optimization (Nelder et al. 1965) for all models. The corresponding best solution (maximum likelihood in case of varieties of $DegT_{50,ts}$, maximum posterior probability in case of $DegT_{50,w}$ and $DegT_{50,sed}$) was used as a starting point for sampling the uncertainty distribution. Sampling was carried out by Markov chain Monte Carlo (MCMC) method based on the classical Metropolis algorithm (Geweke et al. 2003). Three parallel chains were run with 100'000 iterations each, out of which 25'000 were dedicated for burn-in. The MCMC algorithm does not imply any special statistical assumptions on the distribution and (in)dependence of parameters, and is therefore ideal to get an unbiased uncertainty estimation. The product of an MCMC sample is a set of different parameter combinations, which – in case of proper convergence – form a representative sample of their underlying statistical distribution. Since none of our half-lives is directly a parameter of the fitted models, they did not show up directly in the MCMC samples. For each record of MCMC sample the corresponding half-lives were calculated based on the actual parameter values in the given record, which led to an equally long statistically representative sample of the half-lives themselves. Uncertainty intervals were determined from these samples.

#### 2.1.1.7 Joint modelling

It was demonstrated by (Honti et al. 2016) that the joint calibration of different water-sediment experiments belonging to the same compound and sediment can increase the identifiability of biotransformation properties of the compound. Improved identifiability and reduced uncertainty is due to the involvement of more data. Here, 'more' is not only meant in a quantitative sense. When different experimental types are treated together, there is also an increase in aspects besides the increase in the sheer amount of data. Different experimental standards grasp different aspects of the compound's behavior in water-sediment systems. While OECD 309 describes biotransformation (or its practical absence) in thin, well-mixed systems, it does not tell anything about behavior in the presence of settled sediment or under anaerobic conditions. Similarly, OECD 308 tells about the disappearance of the parent compound in different compartments, but it does not give solid evidence about the place (water? sediment? thin aerobic layer of sediment?) and mechanisms of biotransformation. Combining experiments is similar to adding more equations (and therefore more information) to a system of unknowns.

The $k'_{bio}$ concept used for estimating $DegT_{50,w}$ and $DegT_{50,sed}$ was actually developed to make coupling of different experimental types possible. With the above outlined $k'_{bio}$ model framework different water-sediment systems can be described with essentially the same terminology and parameters. Coupling is actually done by forcing the models for the individual experimental systems to share certain parameters. With the coupling we assume that

- the model framework is universally valid for all involved experimental types,
- model parameters and processes mean the same in all systems,
- and therefore identifying a certain process in one experimental system allows us to extrapolate to the corresponding (yet potentially not directly observable) parts of another experimental system.

The coupling of OECD 308 and OECD 309 experiments in practice was done by forcing $k'_{bio}$, $K_d$, and $dK_d$ to be the same in both systems. The relevant model equations suggest that in this case a pelagic OECD 309 system is identical to the water column of the OECD 308, when TOC and DOC concentrations are similar (except for qualitative differences in biomass). Thus, actual biotransformation in water can be learnt from pelagic OECD 309 data and thereby the identifiability of phase-transfer and sediment biotransformation processes in the OECD 308 may increase significantly. In this project the cases with an OECD 308 and a non-pelagic OECD 309 belonging to the same compound and sediment was limited. To increase the quantity of data available for joint modelling, we allowed coupling of OECD 308 to pelagic OECD 309 from a different water body. Thereby we assumed that the small amount of degrader biomass in a pelagic OECD 309 may compensate for the potentially large qualitative differences due to the different source ecosystem.

#### 2.1.1.8 Transformation product case study

Besides characterising the robustness of $DegT_{50,w}$ and $DegT_{50,sed}$ for the parent compound, we made a small feasibility study on deriving the same indicators for transformation products. Two compounds (API8, API13) were selected based on having fast biotransformation from parent to transformation product, and one dominant transformation product, which was again biotransformed relatively fast.

Thus, in the selected cases, there was a clear signal of the transformation product's time dynamics and the dominance of a single product prevented the occurrence of complicated transformation patterns, like concurrent or serial transformation pathways.

For these selected experiments, we analysed $DegT_{50,w}$ and $DegT_{50,sed}$ for the transformation product pool too (biotransformation of TP pool was always considered in the model, there was no need to perform additional calibration or uncertainty analysis). Calculation of these compartment-specific half-lives followed the same equations as for the parent compound, albeit with $K_{oc}$ and $k'_{bio}$ of the transformation products.

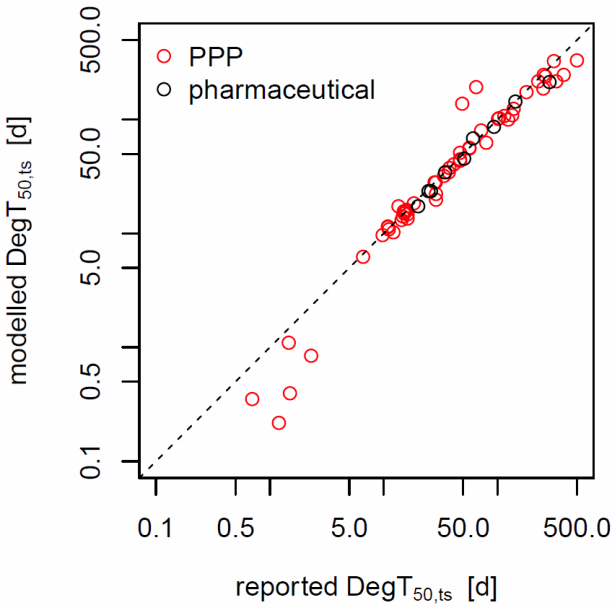
## 2.1.2 Results

### 2.1.2.1 $DegT_{50,ts}$

According to common sense, $DegT_{50,ts}$ should be the most robust among degradation half-lives, because it derives from directly observable and clearly identifiable data (the signal from the parent compound must be clear in all such experiments) and requires fitting very simple models (Honti et al. 2015). Its perceived robustness is reflected in the current assessment practice too: single best values of $DegT_{50,ts}$ are usually reported and compared to strict cutoff values during persistence assessment. However, as both observation data and models are likely to contain errors, $DegT_{50,ts}$ also has its own uncertainty, which is surprisingly not considered in practice.

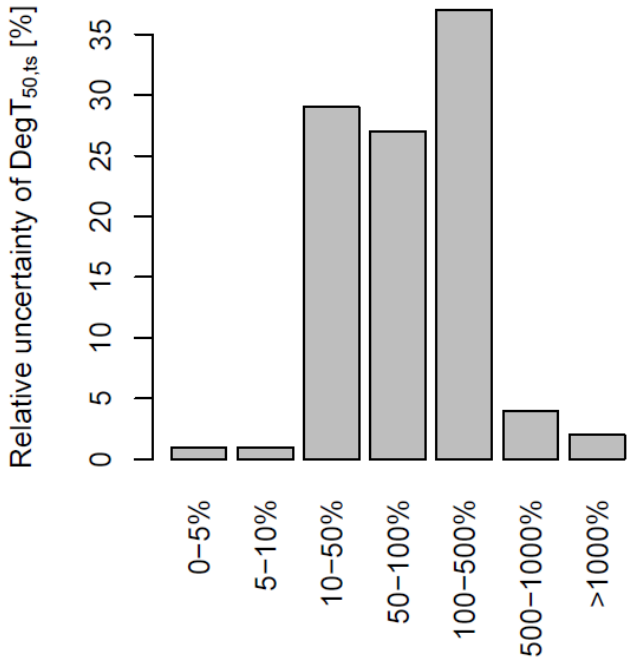
As a first step, newly calculated maximum likelihood $DegT_{50,ts}$ values were compared to the values reported in the dossiers. There was generally a good agreement between the reported and calculated values (Figure 3), so both reported best $DegT_{50,ts}$ values and the newly implemented calculation algorithm could be considered reliable.

Figure 3: Comparing $DegT_{50,ts}$ values from the reports and mean $DegT_{50,ts}$ from our calculation. There is generally a good agreement between the reported and calculated values. The outlier pair above 50 d belong to PE12, having strongly atypical degradation kinetics and consequently very high uncertainty for modelled $DegT_{50,ts}$.




According to the results from the examined cases, the uncertainty of $DegT_{50,ts}$ is not negligible. Since $DegT_{50,ts}$ values vary over orders of magnitude, we analysed their relative uncertainty, that is the width of the 95% confidence interval divided by the mean. Almost half (43%) of the experiments had above 100% relative uncertainty (i.e., a factor of 2) using the SFO model, and 6% of the cases had above 500% relative uncertainty (Figure 4).

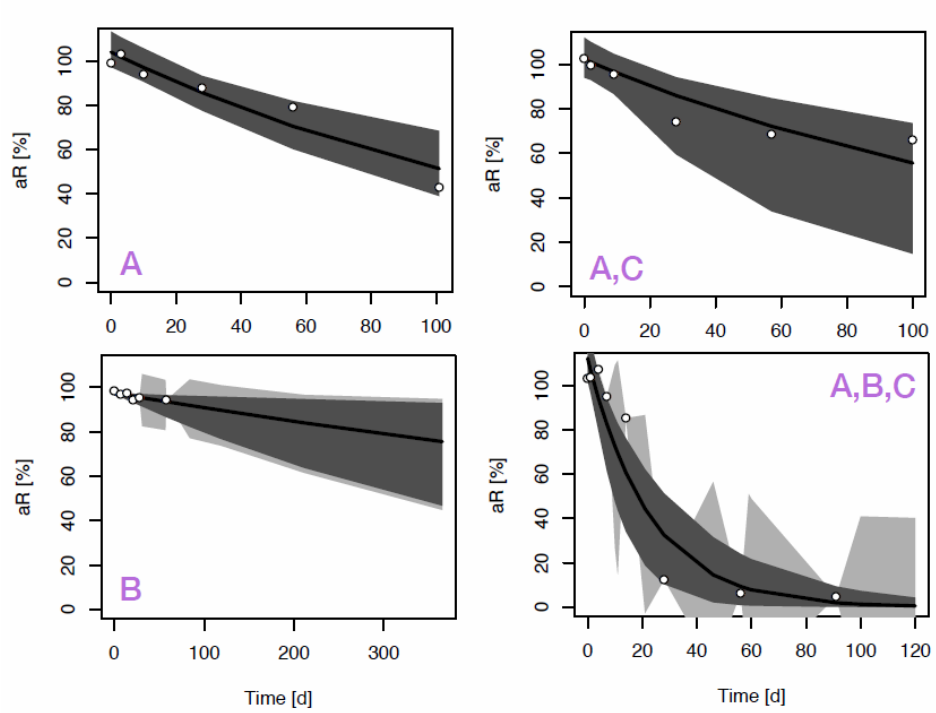
Figure 4: Relative uncertainty of $DegT_{50,ts}$. Relative uncertainty is the width of the 95% confidence interval divided by the mean.




Although these uncertainties are still considerably lower than encountered for the compartment-specific half-lives (see Chapter 2.1.2.2), these occasionally high uncertainties of above 500 % still reveal the theoretical problems of using a solid persistence criteria when classifying the compound based on persistence. A relative uncertainty of 100% would mean that classification around the 40 day water cutoff value would become uncertain if the $DegT_{50,ts}$ was in the interval of 27-80 days, which makes the application of a single cutoff value difficult. A compound classified as non-P based on a $DegT_{50,ts}$ of 27 days, assuming that this value might deviate +50% (=13.5 days when the uncertainty interval is symmetrical) with significant probability, can happen to have a true $DegT_{50,ts}$ just above 40 days. Similarly a persistent compound with $DegT_{50,ts}$ of 80 days may still have the chance to have a true $DegT_{50,ts}$ just below 40 days. The presence of non-negligible uncertainty means that for most compounds the $DegT_{50,ts}$ being below the corresponding cutoff value will have a probability that is neither 0% nor 100%.

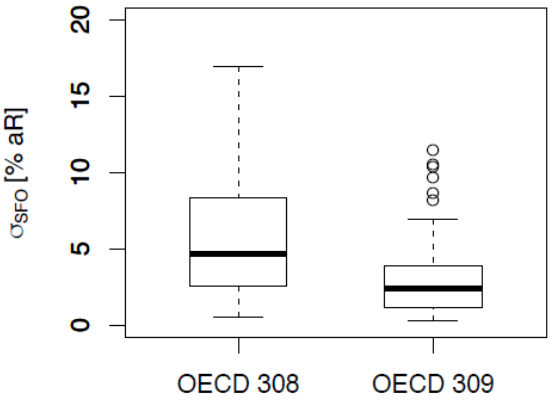
The uncertainty of $DegT_{50,ts}$ originates from different sources (Figure 5). One of them can be the measurement noise (the random deviation of observation points around the smooth degradation curve), which increases the range of uncertainty of the kinetic fitting. As introduced in section 2.1.1.6, high noise levels in the data extend the set of model fits having equivalent total error and therefore widens the uncertainty interval of model parameters. Another source of uncertainty are sparse or missing data points. Model parameters can be identified well when observation points deliver an unambiguous evidence about the shape of the degradation curve. Parameter uncertainty grows if measurements ended prematurely, that is before observing significant degradation, or crucial parts of the decay curve do not have any observation points. A final uncertainty factor is the adequacy of the applied model to describe the observed degradation curve. In many cases the otherwise smooth and well-sampled degradation curves do not follow any of the model functions, which leads again to high fit error and increased uncertainty.

Figure 5: Sources of uncertainty of $DegT_{50,ts}$ demonstrated on some experiments with uncertain fit: high measurement noise (A), sparse or missing data points or premature termination of measurements (B) and atypical degradation kinetics (C). Dark shading: 95% uncertainty interval; points: observations; line: best model fit. Experiments (from top left, clockwise): API2_ae_s14_308; API4_ae_s35_308; PE1_ae_s19_308; PE5_ae_s13_309.



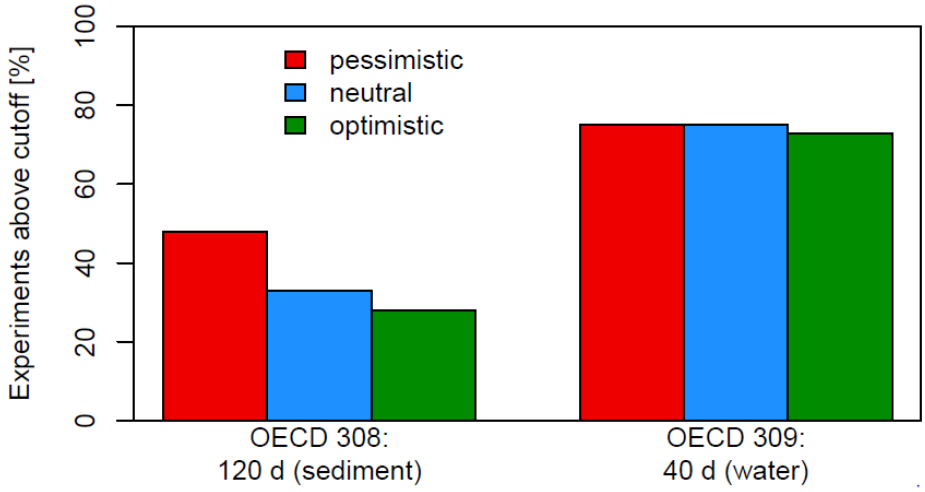

Mean model errors for SFO fits of OECD 309 data were about half the mean errors for SFO fits of OECD 308 data (Figure 6). These results confirm that OECD 309 experiments are conceptually simpler, leading to lower relative uncertainty of $DegT_{50,ts}$ estimated in comparison to OECD 308 experiments. The inferior performance of the SFO model in OECD 308 systems can be related to the fact that total residues in a two-phase system with spiking into one phase only are theoretically not expected to follow SFO at all. Such concerns do not apply to OECD 309, as it's fully mixed. In practice OECD 308 experiments often show time courses that closely resemble first-order kinetics, but this is rather accidental due to partitioning into one phase or measurement error levels hiding the fine details of the degradation curve.

Figure 6: Model error levels for SFO ($\sigma_{SFO}$) in OECD 309 and in OECD 308 tests for all compounds. Mean model error for OECD 309 data was about half of the mean model error for OECD 308 data.




This difference had its consequences for the uncertainty of classification as well. Persistence values derived from OECD 309 led to less uncertainty in classification compared to persistence values derived from OECD 308 (Figure 7). Thus, while it made a large difference if the persistence outcomes derived from OECD 308 experiments were classified in an optimistic, pessimistic or neutral manner (based on the lower uncertainty limit, the higher uncertainty limit or the mean of $DegT_{50,ts}$, respectively), this effect was almost completely missing when the same analysis was done with the persistence outcomes derived from OECD 309. This was partly because there was no significant degradation in most OECD 309 experiments anyway, and hence most compounds fell well above the 40 d persistence criterion anyway. Thus even the uncertainty of $DegT_{50,ts}$ could not alter their classification.

Figure 7 : Percent of experiments above persistence criteria in OECD 309 and OECD 308 tests. Classification: optimistic, pessimistic or neutral (based on the lower uncertainty limit, the higher uncertainty limit or the mean of $DegT_{50,ts}$, respectively). Persistence criteria used: 40 days in water, 120 days in sediment.




The FOCUS Guideline (FOCUS 2014) allows fitting multiple kinetic models to determine $DegT_{50,ts}$. The guideline instructs to have preference for the simple SFO model unless certain statistical criteria are violated by the insufficient model fit. In the examined dataset the simple and robust SFO model fulfilled the chi-square ($\chi^2$) criterion in 111 out of 112 cases (99%). Moreover SFO was not only a sufficiently good solution but even statistically the best in terms of the Akaike Information Criterion (Akaike

1974) in 45% of the cases. Given the complexity and the related stability problems of the more complicated DFOP and HS models, in accordance with section R.7.9.4.1 of (ECHA 2016a) we suggest to use the SFO model exclusively to calculate $DegT_{50,ts}$.

**Conclusion**

From the investigated kinetic models, SFO was the most robust way to calculate $DegT_{50,ts}$, giving sufficiently good fit to the observed data in all but one case. The more complicated DFOP, HS, and FOMC models are over-parameterised for most experiments. They are therefore numerically unstable for scarce data or outside the observed data range.

Even when fitted with the SFO model, $DegT_{50,ts}$ values had a significant uncertainty, which is not acknowledged in current assessment practice. For most cases, the relative uncertainty around the mean exceeded 20%. For about half of the cases it even exceeded 100%. This leads to an incompatibility of the uncertain $DegT_{50,ts}$ and the presently used rigid persistence criteria. It is recommended that, in the future, uncertainty assessment for $DegT_{50,ts}$ should become an integral part of persistence assessment and persistence classification procedures should be able to account for the uncertainty in persistence indicator values. This is even more so since typically differing $DegT_{50,ts}$ values obtained from different water-sediment systems are available, representing natural variability on top of indicator value uncertainty.

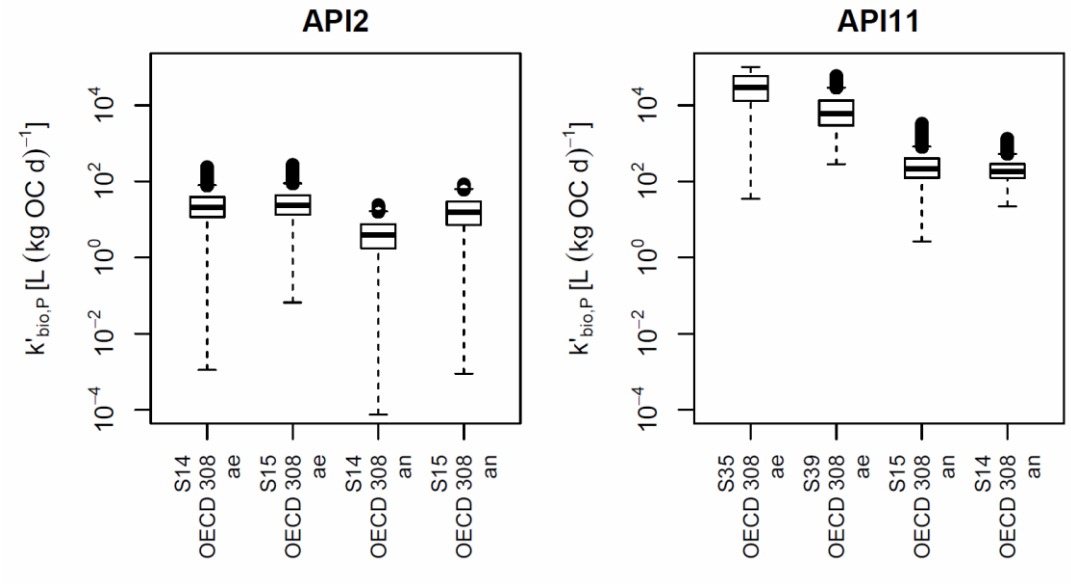
One way to handle the inherent uncertainty and variability of $DegT_{50,ts}$ would be to assign an acceptable probability to the persistence criterion. In many cases, $DegT_{50,ts}$ does not clearly fall below or above the persistence value, i.e., it has certain non-negligible probabilities of being on either side. Classification of these compounds is only possible when this probabilistic aspect is properly handled, for example by defining a small yet acceptable probability for $DegT_{50,ts}$ values to lie above the persistence criterion.

#### 2.1.2.2 $DegT_{50,w}$ and $DegT_{50,sed}$ of individual experiments

The second-order biotransformation rate constant $k'_{bio}$ turned out to be very uncertain in all cases. Accordingly, compartment-specific degradation half-lives were uncertain too. This is in line with the findings of (Honti et al. 2015), who found that the uncertainty of $DegT_{50,w}$ and $DegT_{50,sed}$ typically exceeded one order of magnitude.

For the same compound $k'_{bio}$ showed high variability between studies, even for the same experimental type. The actual sediment used and the redox conditions had a high influence on the derived $k'_{bio}$ values (Figure 8).

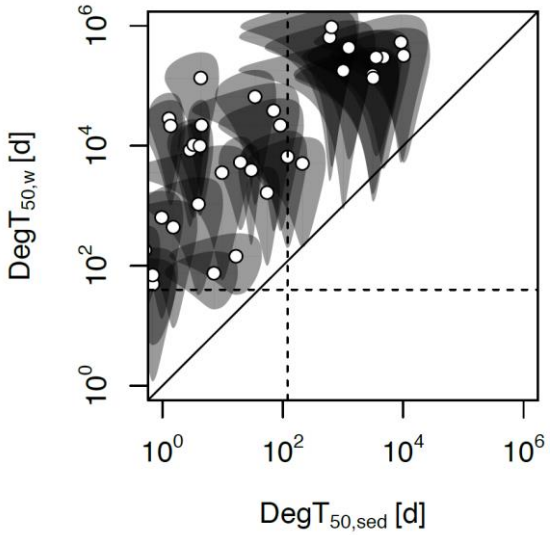
Figure 8: Examples for the high variability of $k'_{bio,P}$ between OECD 308 studies (compounds: API2 and API11). Notation: $k'_{bio,P}$ is the second-order biotransformation rate constant, S refers to sediment, ae and an refer to the test system's redox conditions (aerobic and anaerobic, respectively).




As a general pattern, $k'_{bio,P}$ was often estimated to be (up to orders of magnitude) higher for OECD 309 experiments, than for OECD 308 despite the inferior degree of degradation observable in OECD 309 systems (Figure 16). This difference could be considered a numerical artifact, because in the absence of significant degrader biomass there was no solid evidence about the value of $k'_{bio}$ in OECD 309, or from another perspective there was no solid evidence against $k'_{bio}$ having arbitrarily high values. Aerobic OECD 308 systems most often had higher $k'_{bio,P}$ compared to anaerobic ones. (When individual experiments are analysed, it is impossible to get the proper aerobic $k'_{bio,P}$ value from an anaerobic experiments since $dk_{aer}$ is unknown, so $k'_{bio,P}$ in anaerobic OECD 308 experiments refers to anaerobic biotransformation, while in aerobic experiments it refers to aerobic biotransformation.) Probability of the uncertain $k'_{bio}$ being higher in the anaerobic system than in the aerobic was only 24% on average. The average ratio between aerobic and anaerobic mean $k'_{bio}$ was 4.14. There was always a significant difference between mean $k'_{bio}$ derived from aerobic and anaerobic OECD 308 systems for the same compound and sediment. P-values of the two-tailed t test ranged between $10^{-9}$ and $10^{-259}$. There was less strong disagreement between experiments belonging to the same compound and redox conditions, but a different sediment ($0 < P < 0.098$). However, there were always exceptions from these patterns, so we could not arrive at a conclusive statement on the variability of $k'_{bio}$ as a function of experimental conditions.

As demonstrated by(Honti et al. 2016), the uncertainty of $k'_{bio}$ maps to uncertainty in $DegT_{50,w}$ and $DegT_{50,sed}$. Both $DegT_{50,w}$ and $DegT_{50,sed}$ are uncertain up to at least 1-2 orders of magnitude, and, following the patterns of $k'_{bio}$, they are different in different sediments and under different redox conditions. As a result of the limited availability of degrader biomass in the water phase and based on the relationship between $k'_{bio}$ and the compartment-specific degradation half-lives (equations (28)-(30)), $DegT_{50,w}$ values were always higher than $DegT_{50,sed}$ values for the same compound by up to several orders of magnitude (Figure 9).

Figure 9: $DegT_{50,w}$ and $DegT_{50,sed}$ values from OECD 308 experiments. $DegT_{50,w}$ was generally found to be larger than $DegT_{50,sed}$ by up to several orders of magnitude. Dots: mean values; clouds: 95% uncertainty interval. Dotted lines indicate the compartment-specific cutoff values, the solid line is the 1:1 line.




Degradation half-lives in the water phase of OECD 308 experiments were in the same range as the half-lives obtained from OECD 309 systems. The $k'_{bio}$ concept ensured that biotransformation in the water phase of OECD 308 was inherently kept low due to the low availability of degrader biomass in the water phase. This is not the case when compartment-specific first-order biotransformation rates are direct parameters of the fitted model as was the case in the model version used in (Honti et al. 2015): due to the weak separability of biotransformation and phase transfer, estimated degradation rate constants in the water phase ($k_{wpm}$) may reach or even exceed degradation rate constants in the sediment phase ($k_{spm}$) even in the absence of hydrolysis or photodegradation. This would suggest a dramatic difference in the compound's biotransformation behaviour in the water phases of OECD 308 and 309, which seems unrealistic.

The fact that properly identified $DegT_{50,w}$ always exceeds $DegT_{50,sed}$ is in interesting contrast with the cutoff values used to identify persistent compounds in these two compartments (40 and 120 days for water and sediment, respectively). However, it is in line, and also explains, the practical experience that for the same set of compounds much more OECD 309 results lead to a persistence classification than OECD 308 results do.

#### 2.1.2.3 Comparison of $DegT_{50,ts}$ and compartment-specific half-lives in individual experiments

Because $DegT_{50,ts}$ is still widely used as a persistence indicator (despite the concerns about its strong dependence on the exact conditions in the experimental system) and it is indeed still less uncertain than the compartment-specific half-lives, we investigated the empirical relationship between $DegT_{50,ts}$, and $DegT_{50,w}$ or $DegT_{50,sed}$, respectively.

It has been hypothesized that for strongly sorbing compounds, almost all relevant processes are likely to take place in the sediment phase of an OECD 308 due to the immediate and full migration of the spiked compound from the water column into the sediment. This suggests that $DegT_{50,ts}$ should roughly equal $DegT_{50,sed}$ in such cases and could therefore legitimately be considered to represent persistence in sediment. Based on the data used in this project, there seems indeed to be a weak positive relation-

ship between $DegT_{50,ts}$ and $DegT_{50,sed}$ for OECD 308 experiments (Figure 10), but it deviates from a simple 1:1 line relationship with no obvious dependence of that deviation on sorption properties (i.e., $K_{oc}$) or substance class. Yet, if the range of compounds is extended to more strongly sorbing compounds by including the PPPs and pharmaceuticals used in the literature study by Honti and Fenner (2015), a convergence of $DegT_{50,sed}/DegT_{50,ts}$ towards unity above a $K_{oc}$ of about 5000 L/kg starts to emerge (Figure 11). So while the theoretically derived $DegT_{50,sed} \approx DegT_{50,ts}$ rule seems to be legitimate for strongly sorbing compounds and can be used to attribute total system degradation to the sediment alone, it is presently difficult to formulate a relationship between $DegT_{50,sed}$ and $DegT_{50,ts}$ for less strongly sorbing substances.

Figure 10: Relation between $DegT_{50,ts}$ and $DegT_{50,sed}$ for OECD 308 experiments. Black circles: pharmaceuticals, red circles: pesticides. Higher circle diameter means higher $K_{oc}$ value (organic carbon-water partition coefficient).

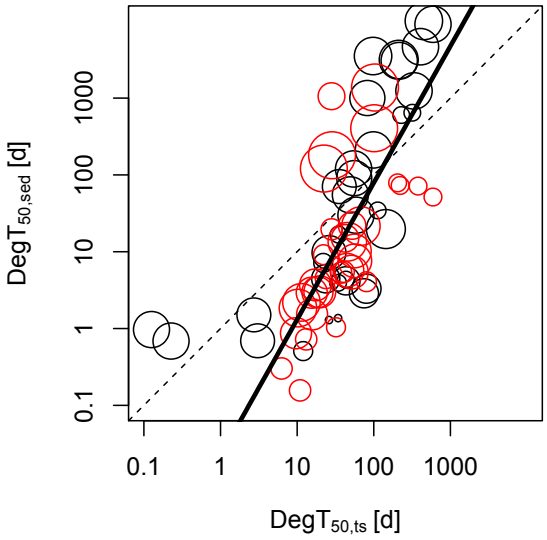
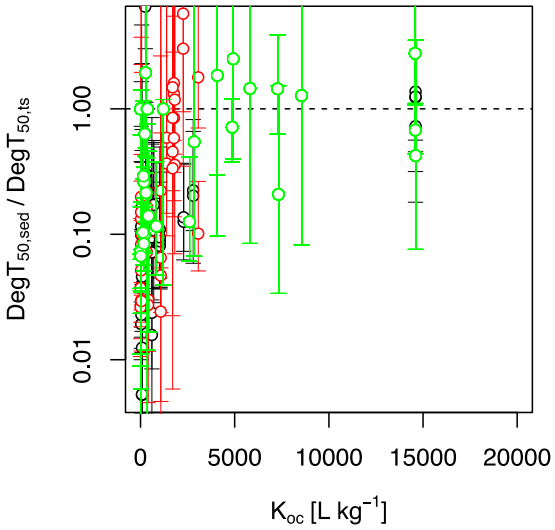
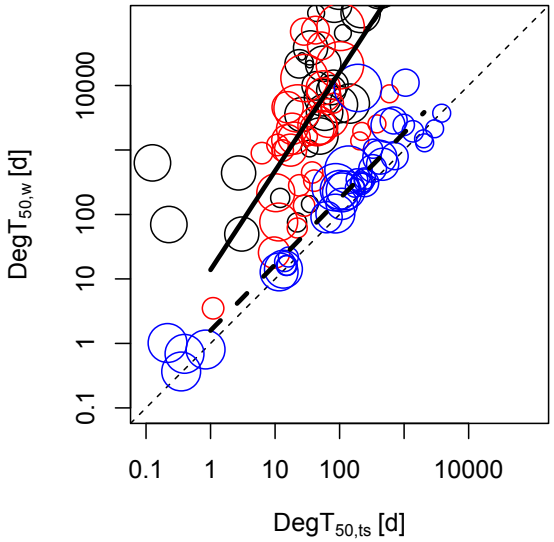



Figure 11: Convergence of $DegT_{50,sed}/DegT_{50,ts}$ towards unity with increasing $K_{oc}$. Black signs: pharmaceuticals from this project, red signs: pesticides from this project, green signs: pharmaceutical and PPP data from (Honti et al. 2015).



A relationships with $DegT_{50,ts}$ can also be discovered for $DegT_{50,w}$ (Figure 12). Since this indicator can be extracted from both OECD 308 and 309 experiments, an interesting distinction can be made. According to the model assumptions, biotransformation can only take place in the water phase of an OECD 309 system, because we consider sorbed fractions to be unavailable. In an OECD 309 system the sorbed fraction is usually negligible unless $K_{oc}$ is very high (above $10^4$-$10^5$ L $kg^{-1}$) due to the limited amount of particulate organic carbon available for sorption. Therefore, $DegT_{50,w}$ is actually very close to $DegT_{50,ts}$ in OECD 309 experiments if no hydro- or photolysis is possible (Figure 12). Obviously, the same does not hold for OECD 308 due to the interactions of the water and sediment phases and therefore the relationships between DegT50,ts and $DegT_{50,w}$ for OECD 308 and 309 are entirely different.

Figure 12: Relation between $DegT_{50,ts}$ and $DegT_{50,w}$ for OECD 308 and for OECD 309 experiments. Black circles: pharmaceuticals (OECD 308), red circles: pesticides (OECD 308), blue circles: pesticides (OECD 309). Higher circle diameter means higher Koc value (organic carbon-water partition coefficient).




#### 2.1.2.4 Transformation product case study

The dominant transformation products of compounds API8 and API13 both were well simulated by our model. We obtained well defined $DegT_{50,sed}$, $DegT_{50,w}$ and $k'_{bio,M}$ values with ranges of relative uncertainty typical of those also observed for parent compounds (Table 6 and Figure 13 and Figure 14). In 3 out of 4 experiments, biotransformation rates were lower for the transformation products than for parent compound. On average, the relative uncertainty ranges of the indicators were somewhat wider for transformation products than for parent compounds (Table 6). Thus, it seems that biotransformation properties of an abundant and dominant transformation product can be extracted from OECD 308 data with about the same accuracy as for the parent compound. However, in most cases, transformation products are diverse and show up in low absolute quantities. For these more typical cases models would need more state variables and parameters due to the differences in $k'_{bio}$ and $K_{oc}$ of the individual transformation products, which would further increase the uncertainty of the extracted half-lives.

Table 6: Comparison of degradation indicators ($DegT_{50,sed}$, $DegT_{50,w}$ and $k'_{bio}$ together with mean and relative uncertainty ) of the parent compounds and their main transformation products in water and in sediment. Relative uncertainty is the width of 95% confidence interval divided by the mean. TP: transformation products, P: parent compound.

| Compound | TP API8 | P API8 | TP API8 | P API8 | TP API13 | P API13 | TP API13 | P API13 |
|---|---|---|---|---|---|---|---|---|
| **Sediment** | **S24** | **S24** | **S17** | **S17** | **S27** | **S27** | **S13** | **S13** |
| **Experiment** | **OECD 308** | | | | | | | |
| **$DegT_{50,sed}$ mean [d]** | **96.8** | **19.8** | **0.6** | **3157** | **129.7** | **7.1** | **41** | **0.5** |
| *rel.unc.* | *1.6* | *1.3* | *2.8* | *2.7* | *2.5* | *2.3* | *1.9* | *1.4* |
| **$DegT_{50,w}$ mean [d]** | **37933** | **5250** | **15.8** | **133451** | **1702** | **75.3** | **16008** | **180.7** |
| *rel.unc.* | *2.7* | *2.6* | *2.9* | *2.9* | *2.5* | *2.1* | *2.7* | *2.5* |
| **$k'_{bio}$ mean [L (kg OC d)$^{-1}$]** | **9.4** | **55.2** | **44732** | **15.7** | **1696** | **22163** | **89.3** | **5353** |
| *rel.unc.* | *1.6* | *1.3* | *2* | *2.7* | *1.9* | *1.4* | *1.5* | *1* |

Figure 13: API13 compound (Case 2: API13_ae_s27_308): parent and transformation product degradation. Top left: $DegT_{50,w}$, top right: $DegT_{50,sed}$, bottom left: $k'_{bio,P}$ or $k'_{bio,M}$. Dashed lines: at 40 days at $DegT_{50,w}$ and at 120 days at $DegT_{50,sed}$. TP: transformation products, P: parent compound.

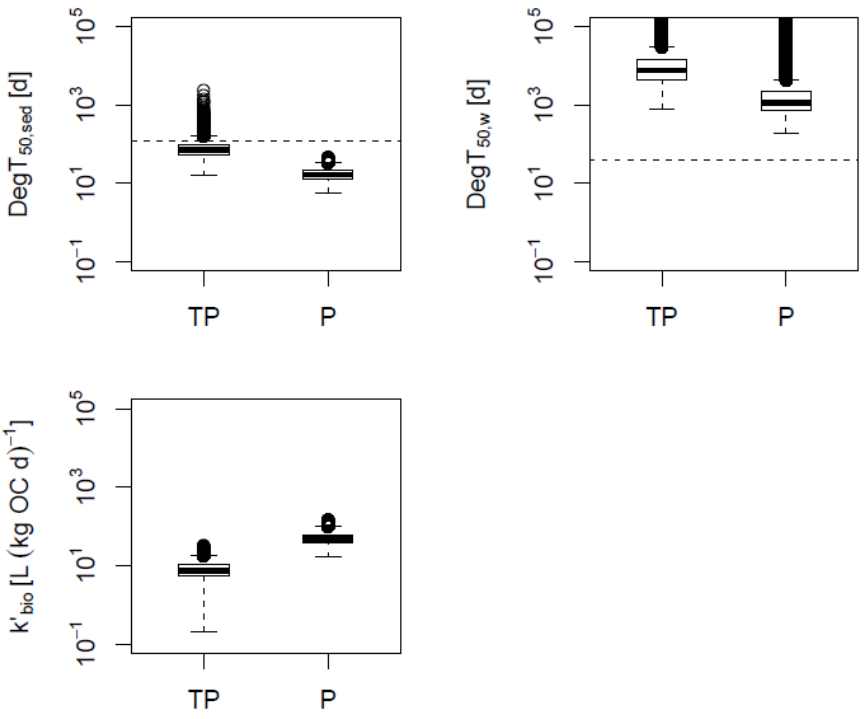
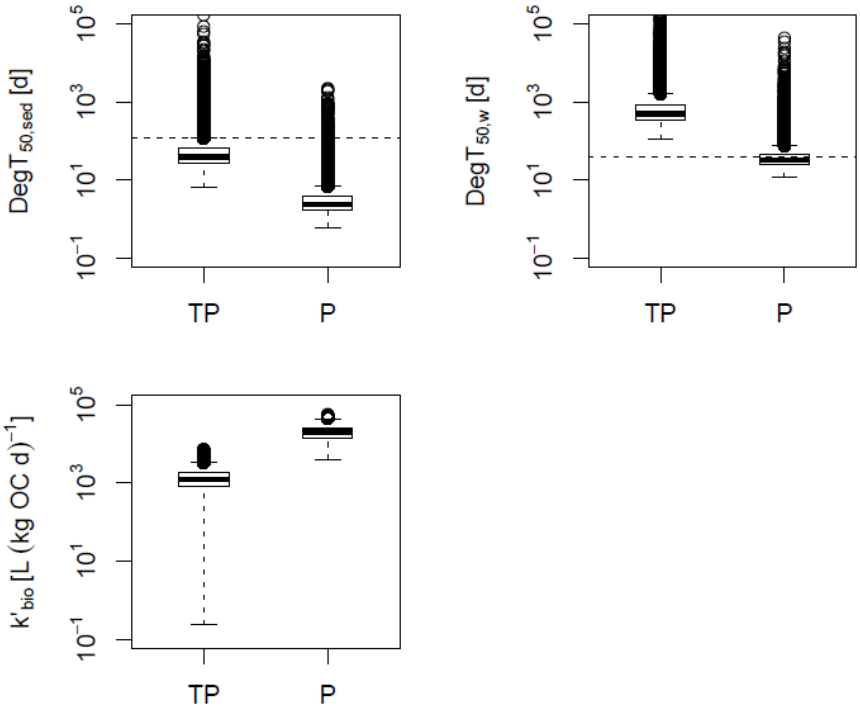


Figure 14: API8 compound (Case 2: API8_ae_s24_308): parent and transformation product degradation. Top left: $DegT_{50,w}$, top right: $DegT_{50,sed}$, bottom left: $k'_{bio,P}$ or $k'_{bio,M}$. Dashed lines: at 40 days at $DegT_{50,w}$ and at 120 days at $DegT_{50,sed}$. TP: transformation products, P: parent compound.



### 2.1.2.5 Joint model of OECD 308 and 309 tests

The joint modeling of OECD 308 and OECD 309 tests belonging to the same compound (and sediment for non-pelagic OECD 309) has again worked properly just like for more experimental varieties in (Honti et al. 2016). The best joint solution is unavoidably suboptimal for individual experiments, but the distance from the individual optima characterizes the tradeoff of connecting the datasets. In our case, joint calibration had comparable error levels as individual calibration (Figure 15 and Figure 16). Hence there was no statistical evidence against connecting the experimental pairs. The lack of problems due to coupling OECD 308 and pelagic OECD 309 from different sites suggests that the qualitative differences in degrader biomass and organic matter were negligible compared to the activity differences between the two systems.

Figure 15: Example for degradation processes in joint fit case of OECD 308 and 309 experiments: common $K_{OC}$, $k'_{bio}$.

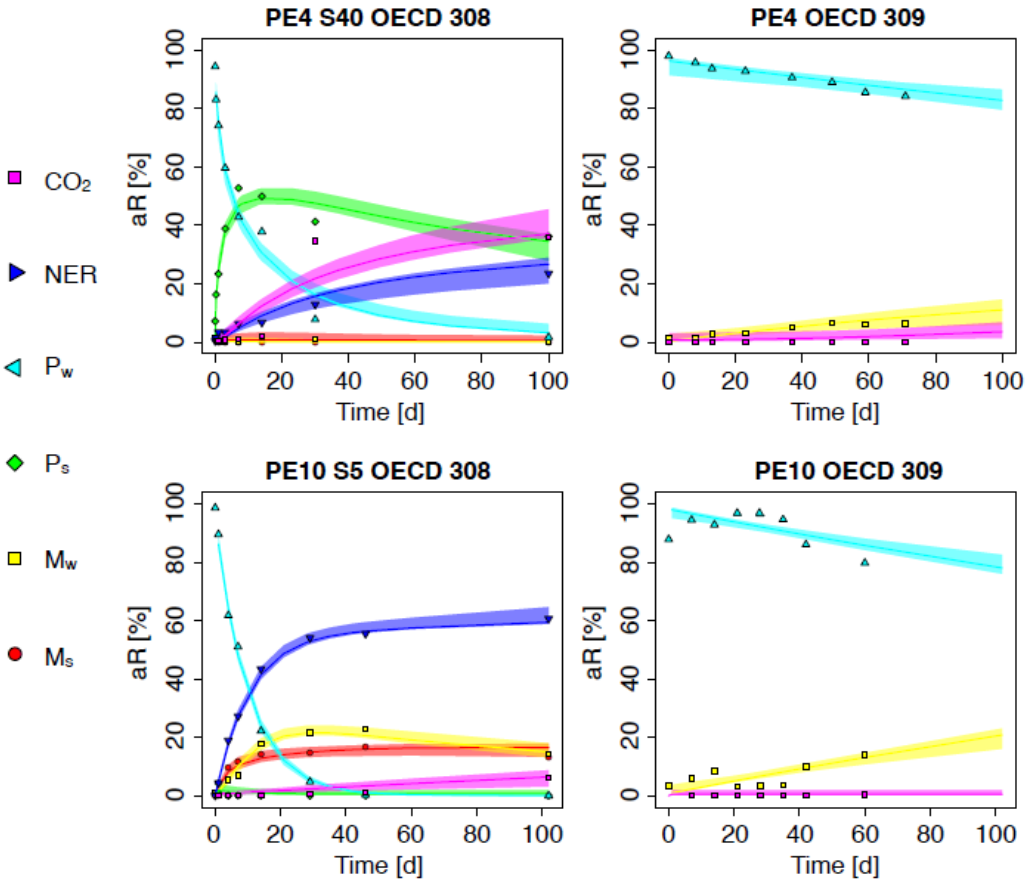
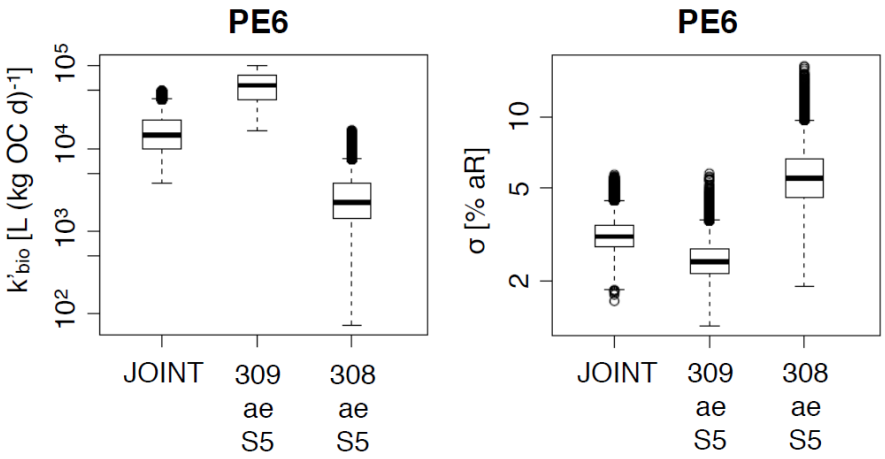


Figure 16: Comparing the joint model and the connecting individual experiments' $k'_{bio,P}$ and error level (σ).




The proof of concept study of connecting experiments (Honti et al. 2016) found that joint calibration narrowed the uncertainty interval of $k'_{bio}$ compared to individual experiments. With the data set used here, this effect was only partial and limited. The relative uncertainty of $k'_{bio,P}$ was in half of the cases smaller for the joint solution than for any of the two individual experiments. In 58% of the cases the joint solution's relative uncertainty was narrower than for the OECD 308. In 92% of the cases the joint $k'_{bio,P}$ was less uncertain than the estimate for OECD 309. The joint $k'_{bio,P}$ was in all cases less uncertain than at least one of the individual experiments. In contrast to (Honti et al. 2016) there was no case when the relative uncertainty of the joint $k'_{bio,P}$ estimate was more than an order of magnitude smaller than for individual experiments. Uncertainty intervals narrowed by only 8% on average, actual narrowing took place between 4 and 16%.

The reason for this limited improvement was due to the serious imparity between the biotransformation-related information content of OECD 308 and 309 experiments. In accordance with practical observations, the lack of significant degradation in most OECD 309 systems seriously impairs the identifiability of $k'_{bio}$ and compartment-specific degradation half-lives from those systems in general. Even in OECD 309 systems with suspended sediment, the typical observed level of degradation is still comparable to the measurement noise (few % of applied radioactivity), and the signal to noise ratio is obviously much worse for pelagic variants. There were a few PPP compounds where both types of OECD 309 results were available. Despite the presence of sediment in one variant, it was difficult to see the excess degradation from the data directly. All of these observations suggest that an OECD 309 system is not likely to dominate over any OECD 308 during joint modeling in determining $k'_{bio}$.

Besides the variability of $k'_{bio}$ estimates, there were systematic differences between the joint and the corresponding individual solutions too. The joint estimate for $k'_{bio}$ usually was in between a smaller estimate belonging to the OECD 308 and a higher estimate belonging to the OECD 309. The bias between the estimates, that is the difference between the mean values, was distributed asymmetrically. The mean of the joint solution fell always closer to the mean of the OECD 308 solution than to the mean of the OECD 309. Again, this could be attributed to the almost negligible information content of OECD 309. Bias was already significant between the joint solution and OECD 308 (−73% to +206% of the joint mean), but practically exploded in relation to OECD 309 (+320% to 11900% of the joint mean).

Based on these findings, we still see a potential in joint modeling of different experimental types to reduce the uncertainty and increase the robustness of persistence indicators. However, OECD 309 systems seem to be inferior, information-poor pairs of the OECD 308 relative to this aspect.

### 2.1.3 Conclusions and recommendations

Based on existing scientific evidence, $DissT_{50,w}$ and $DissT_{50,sed}$ mix up biotransformation and phase transfer. These indicators are therefore not recommended for persistence classification or exposure assessment. Instead, compartment-specific half-lives are conceptually superior by indicating biotransformation separately. The derivation of $DegT_{50,w}$ and $DegT_{50,sed}$, however, requires sophisticated modeling.

Modeling based on $k'_{bio}$ allows estimating $DegT_{50,w}$ and $DegT_{50,sed}$ in a rather universal way (in terms of water-sediment systems). However, just like $DegT_{50,w}$ and $DegT_{50,sed}$ derived from first-order biotransformation rates, $DegT_{50,w}$ and $DegT_{50,sed}$ derived from $k'_{bio}$ are still very uncertain. The joint calibration of different experimental types at once via shared model parameters reduces the uncertainty of $k'_{bio}$, $DegT_{50,w}$ and $DegT_{50,sed}$, but this reduction of uncertainty is limited due to the limited information content of OECD 309 data from pelagic systems.

$DegT_{50,w}$ and $DegT_{50,sed}$ of abundant and dominant transformation products can be extracted with comparable accuracy as for the parent compound, yet the required abundance and dominance of a single transformation product limits such exercises to a small set of compounds.

The uncertainty inherent in all persistence indicators derived from concentration-time series data highlights a serious weakness of the present persistence assessment framework. Comparison of optimal values of persistence indicators to cutoff values ignores the significant uncertainty, yet a robust and accepted way of acknowledging this omnipresent uncertainty is not defined in any regulatory framework relevant for persistence assessment.

The above outlined issues and results highlight that significant improvement could be made to the assessment of a compounds' degradation process with relatively small efforts. We suggest minor modifications in the structure and the content of the reports, and also in the assessment method.

#### 2.1.3.1 Reports

Based on the experience gathered during data collection (see 2.1.1.2) and quality control, we propose minor amendments to the current reporting requirements of OECD 308 and OECD 309 studies. Considering the present variety of reporting formats and acknowledging the existence of aspects other than persistence assessment, we suggest that reporting a small set of crucial data should be made obligatory in a predefined format (Table 7).

Table 7: Suggestions for obligatory supplemental information or metadata in OECD 308 and 309 studies

| Piece of Information | Important |
|---|---|
| (an)aerobic conditions | in evaluating the biotransformation process |
| Name of sampling site | as supplementary information |
| Concentration of the compound in the flask | as supplementary information |
| Temperature | as supplementary information |
| Sediment organic carbon content [%] | in calculation of the degradation process |
| Wet sediment height [cm] (OECD 308) | in sediment's porosity calculation, in calculation of the degradation process |
| Sediment dry weight [g] (OECD 308) | in sediment's porosity calculation |
| Sediment wet weight [g] (OECD 308) | in sediment's porosity calculation |
| Flask geometry | in sediment's porosity calculation, in calculation of the degradation process |
| Wet sediment volume [$cm^3$] (OECD 308) | in sediment's porosity calculation |
| Water column height [cm] | in sediment's porosity calculation, in calculation of the degradation process |
| Water column volume [ml] | in sediment's porosity calculation, in calculation of the degradation process |
| Porosity of wet sediment (OECD 308) | in calculation of the degradation process |
| Total organic carbon concentration in water (TOC) [mg OC $L^{-1}$] | in calculation of the degradation process |
| Dissolved organic carbon concentration in water (DOC) [mg OC $L^{-1}$] | in calculation of the degradation process |
| Sediment–water partitioning coefficient ($K_d$) [L $kg^{-1}$] | in calculation of the degradation process - require $K_d$ determined from the actual sediment |
| Total suspended sediment content (TSS) [kg $L^{-1}$] (OECD 309) | in calculation of the degradation process |
| Possibility of hydrolysis. If hydrolysis is possible, what is an estimated rate of it [$d^{-1}$]? | in separating hydrolysis from biotransformation |

Besides these crucial metadata, a precompiled residue table (example: Table 5 in section 2.1.1.2) would help the assessment of data quality in terms of consolidated recovery rates and material balances.

#### 2.1.3.2 Degradation half-lives

Currently, in OECD 308 and 309 reports, the parent compound's $DegT_{50,ts}$, $DissT_{50,w}$ and $DissT_{50,sed}$ (if applicable) are given. These parameters form the basis of the persistence assessment process. There is scientific consensus that $DissT_{50}$ values of OECD 308 are strongly system-specific and do not (only) reflect degradation due to a heavy influence of diffusion into the (Mechteld et al. 2016). On one hand, this prevents their meaningful application in exposure models. On the other hand, this system-dependence is often so strong that the comparison of compounds based on their behaviour in the same type of experimental system may become biased due to the small allowed variations in experimental boundary conditions (Honti and Fenner, 2015). Therefore, the obligatory calculation and reporting of $DegT_{50,w}$ and $DegT_{50,sed}$ would provide a more relevant and complete picture of the actual persistence of the studied compound.

We suggest to make the calculation of the uncertainty of all reported half-lives obligatory. For the calculation of compartment-specific degradation half-lives, we suggest to use the model framework based on $k'_{bio}$.

The joint modelling of OECD 308 and 309 systems could be used to reduce the uncertainty of the compartment-specific half-lives, yet the pelagic OECD 309 was found to provide only weak information about biotransformation itself. In the available dataset the difference between the pelagic and suspended sediment versions of OECD 309 was smaller than the typical measurement accuracy. It should therefore be discussed in the future whether performing a modified OECD 309 with higher suspended sediment concentrations as suggested by Shresta et al. (2016) (Shrestha et al. 2016) should be considered for the purpose of confidently deriving $k'_{bio}$.

# 3. Factors influencing persistence in laboratory and field systems

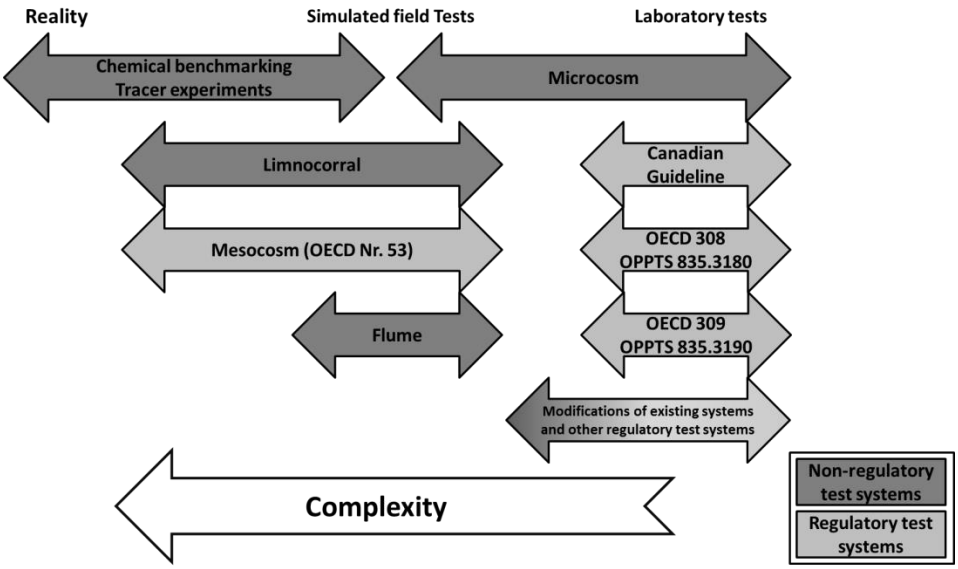
This Chapter presents a literature review of the major factors influencing biotransformation of chemicals in surface water bodies, particularly river systems, and how these are reflected in different laboratory-based test systems. It also provides a comparison of half-life data measured in different laboratory test systems. It ends with a discussion of how well the laboratory-based tests represent the actual conditions in natural river systems.

## 3.1 Existing experimental approaches to quantify persistence

A review of the existing literature on experimental systems to determine biotransformation or biodegradation of chemicals in water-sediment systems was conducted. The review focused on standardized test systems that are mainly used for regulatory purposes, but also reviewed the scientific literature for reports on strongly deviating or complementary experimental approaches. The literature review was restricted to covering studies published in the years 2005-2015 with the exception of some earlier key studies that are also included.

The overview of existing test systems showed that they can be grouped into tests carried out directly in the natural environment, partially laboratory-based, mesocosm-type tests, and fully laboratory-based test systems. Figure 17 shows an attempt to structure the variety of experimental approaches. Detailed descriptions of each system are given in Chapter 3.2. Laboratory-based tests are the major type of tests used in a regulatory context due to their better reproducibility and lower costs compared to field-based experiments conducted in real-world aquatic systems. The latter more realistically capture the influence of environmental factors on degradation, but lead to high operational costs, reduced replicability and comparability. In the context of this review, field-based experiments are considered benchmark systems against which the validity of laboratory-based tests is assessed.

Figure 17: Structured overview over existing test systems to quantify persistence in water-sediment systems




All tests carried out to determine biotransformation or biodegradation of chemicals in water-sediment systems are meant to provide some kind of persistence or half-life information, amongst other things.

This typically requires that the disappearance of the chemical of interest or the appearance of relevant products is measured over time. From the resulting concentration-time profiles, different persistence indicators can be derived as discussed in detail in Chapter 2. Test systems may differ as to what concentrations are quantified (e.g., parent compound disappearance, $CO_2$ formation, total system versus compartment-specific concentrations). This can lead to qualitatively different meanings of the persistence indicators derived from different tests.

In the following, regulatory test systems and alternative approaches described in the literature are compiled and discussed with respect to (i) their field of application, (ii) their specific objectives, (iii) their experimental setup, and (iv) persistence endpoints derived from the test outcomes.

## 3.2 Test system descriptions

### 3.2.1 Regulatory test systems

#### 3.2.1.1 OECD 308

*Reference:* Aerobic and Anaerobic Transformation in Aquatic Sediment Systems, OECD guideline for the testing of chemicals, 2002 (OECD 2002)

*Field of application:* Water-sediment biodegradation studies in aquatic systems

*Objective:* Determination of degradation and transformation at the water-sediment interface of lentic systems under aerobic and anaerobic conditions

*Test system setup:* The test simulates a stagnant system with a settled sediment layer. The test should be run under aerobic and, optionally, also under anaerobic conditions. Two sediments are normally used. The two sediments selected should differ with respect to organic carbon content and texture. For strictly anaerobic studies, the two sediments (including their associated waters) should be sampled from the anaerobic zones of surface water bodies. The test should be performed in the incubation apparatus with a water-sediment volume ratio between 3:1 and 4:1, and a sediment layer of 2.5 cm (± 0.5 cm). The test should be performed in the dark at a constant temperature in the range of 10 to 30°C. The duration of the experiment should normally not exceed 100 days. The test chemical should be applied as an aqueous solution into the water phase. Throughout the test, the sediment should be disturbed as little as possible. The use of $^{14}C$-labelling is recommended to measure rates of transformation and mineralization.

*Derivation of persistence endpoints:* Typically, disappearance half-lives are derived for both the water and sediment compartment individually ($DT_{50w}$, $DT_{50sed}$) and the total system ($DT_{50ts}$) (FOCUS 2006). These indicators can be derived directly from the compartment-specific concentration-time courses and the total concentration-time course, respectively. It is also recommended that compartment-specific degradation half-lives be derived ($DegT_{50,w}$, $DegT_{50,sed}$), which, however, requires inverse modeling (FOCUS 2006, Honti et al. 2015).

#### 3.2.1.2 OPPTS 835.3180

*Reference:* Water-sediment Microcosm Biodegradation Test, United States Environmental Protection Agency / 1998 (EPA 1998)

*Field of application:* Water-sediment biodegradation studies in aquatic systems

*Objective:* To establish criteria of minimum acceptability for the development of water-sediment microcosms for use in biodegradation studies

*Test system setup:* The test guideline does not include any specific microcosm design. Handling and operation should be compound- and site-specific. The described microcosms contain sediment and

water that have been collected from test sites in a manner that maintains the physical and biological integrity of the ecosystem under study. Physical parameters such as lighting, mixing, and temperature may be controlled to simulate the environmental conditions of the site from which water and sediment samples are collected. Sediment samples should be checked for chemical contamination, which should be avoided as much as possible. If needed, the contamination history should be considered when choosing the initial test substance concentrations. Test compounds are introduced into the system either as a single dose or by continuous dosing over the entire test duration. The microcosms should be sampled on a periodic basis, and the water, sediment and off-gases analyzed for disappearance of the parent compound and, if feasible, appearance of transformation products. Use of sterile control microcosms permits determination of the relative importance of biotic and abiotic processes in the fate of a test compound. The duration of microcosm testing should be limited to 60 days or less, unless specific circumstances and microcosm function warrant longer operation. Microcosms may be maintained in either flow-through or static-renewal modes.

*Derivation of persistence endpoints:* The guideline only specifies that a $DT_{50w}$ value should be derived, i.e., a disappearance half-life of the test substance from the water column.

#### 3.2.1.3 OECD 309 and OPPTS 835.3190

*Reference:* Aerobic Mineralization in Surface Water – Simulation Biodegradation Test, OECD guideline for the testing of chemicals, 2004 (OECD 2004)

*Field of application:* Aerobic biodegradation studies in surface waters

*Objective:* Determination of mineralization (and optionally transformation) of test substances in surface waters

*Test system setup:* The test is performed as laboratory shake flask batch test by incubating the test substance with either surface water only or surface water amended with suspended solids/sediment. The first option can be used for simulating biodegradation in surface water free of coarse particles ("pelagic test"), whereas the second option is simulating turbid surface waters which, for example, might exist near a water-sediment interface ("suspended sediment test"). At least two different concentrations of test substance should be used in order to determine the degradation kinetics. For the optional suspended sediment test, surface sediment is added to the flasks containing natural water. Incubation should take place in the dark (preferred) or in diffuse light at a controlled (±2°C) temperature, which may be the field temperature or a standard temperature of 20-25°C. Agitation by means of continuous shaking or stirring must be provided to maintain particles and microorganisms in suspension. The duration of the test should normally not exceed 60 days. If possible, the test chemical should be applied as an aqueous solution into the water phase. The use of $^{14}C$-labelling is recommended to measure rates of transformation and mineralization.

*Derivation of persistence endpoints:* Typically, total system degradation and mineralization half-lives are derived from the data ($DegT_{50,ts}$). Whereas the former is derived from a plot of the parent substance concentration versus time, the latter is derived from total $^{14}C$ remaining in the system versus time.

#### 3.2.1.4 OECD 53 (Micro- and mesocosms)

*Reference*: Guidance document on simulated freshwater lentic field tests (outdoor microcosms and mesocosms), 2006, (OECD 2006)

*Field of application:* Simulation of lentic fresh water systems

*Objective:* To obtain more knowledge about the ecological relevance of effects identified in laboratory studies of lentic systems. Since the substances' fate in these system is an integral part of understanding the observed effects, the data collected may also lend itself to study the substances' primary transformation in the aquatic environment under natural or semi-natural conditions.

*Test system setup:* Outdoor meso- or microcosm studies can be performed in artificial tanks or ponds or by enclosing parts of existing ecosystems. The meso- or microcosms can be constructed from any natural substrate or inert material. A size of 1 to 20 $m^3$ is usually regarded as appropriate for outdoor meso- or microcosm studies. Sediments should always be included in the test systems. The utilized water should originate from the zone where the sediment and its organisms were collected. The test includes macrophytes, invertebrates and fish. Application of the test substance can be conducted either by direct addition of the chemical to water or by simulating the actual route of exposure.

In addition to the stated OECD guideline multiple other researchers present mesocosm and microcosm studies investigating the environmental fate of chemicals. The experiments vary in size and conditions. Non-regulatory microcosm studies on chemical fate have been conducted by, amongst others, (Lam et al. 2004, Hand et al. 2010, Boonstra et al. 2011, Caracciolo et al. 2012). Additional mesocosm studies have been presented, amongst others, in (Bromilow et al. 2006, Sanderson et al. 2007)

*Derivation of persistence endpoints:* No recommendations on the derivation of persistence endpoints are given since this is not the main purpose of the guideline. However, since quantification of exposure concentrations is required anyway, these data could be used to derive transformation half-lives for the parent compound. Depending on the type of experimental system used, this may require correction for volatilization and continued outflow of the test substance, however.

#### 3.2.1.5 Canadian Guideline for Determining Environmental Chemistry and Fate of Pesticides

*Reference:* Guidelines for Determining Environmental Chemistry and Fate of Pesticides, 1987 (AND et al. 1987)

*Field of application:* Registration of pest control products

*Objective:* Determination of data to derive the fate of pesticides in the aquatic environment (Chapters 6.2.C.2 and 6.3.B)

*Test system setup:* (i) Aquatic biotransformation in the laboratory (Chapter 6.2.C.2): Rates of degradation of the parent compound should be tested under anaerobic and aerobic conditions at two constant temperatures, one in a lower and one in a higher temperature range. Aerobic incubations should be carried out under a standard lighting regime, e.g., 16 hrs. light, 8 hrs. dark, using fluorescent lights of the type recommended for plant cultivation. Anaerobic incubations should be carried out in the dark. Aerobic degradation should be determined in unfiltered natural water held under static conditions, or aerated by shaking or bubbling with air. Where physicochemical properties of the pesticide (e.g., adsorption/desorption parameters) suggest that sediment will be a major sink for the pesticide, degradation should be studied in water-sediment systems rather than in unfiltered water. Anaerobic incubations in water-sediment systems should be carried out concurrently with aerobic incubations. (ii) Dissipation and accumulation in the field (Chapter 6.3.B): Small-scale aquatic field studies in natural or artificial small ponds or enclosures (1-5 $m^3$) and with little or no inflow or outflow. Compartments of aquatic environments (e.g., water, sediment, biota) that are critical to the fate of a particular substances must be sampled thoroughly. Application should follow a "worst-case" scenario, e.g., inadvertent direct spray by aircraft or spray drift from adjacent field applications. Use of radiolabeled pesticide in small-scale studies may be considered as a means of estimating a mass balance. Sampling should be carried out prior to treatment, immediately after treatment and at increasing intervals between samplings (daily, weekly, monthly) depending on estimates of field dissipation from laboratory data.

*Derivation of persistence endpoints:* No recommendations on the derivation of persistence endpoints are given.

#### 3.2.1.6 American Pesticide Assessment Guidelines

*Reference:* Pesticide Assessment Guidelines,*(USEPA 1982)*

*Field of application:* Performance of environmental fate testing to support registration of pesticides

*Objective:* Determination of the nature and extent of formation of pesticide residues in water and in hydrosoil (anaerobic). Determine the effects on a pesticide of exposure to aerobic conditions in water or sediment during the period of dispersal of the pesticide throughout the aquatic environment (aerobic).

*Test system setup:* The described test can either be conducted under anaerobic or aerobic conditions. The water and sediment used for the test should be obtained from the study site of interest (flooded soil may be used as a substitute to sediment for the anaerobic test) and the test is run as a layered system. The test substance should be applied at a rate sufficient to permit measuring the disappearance of the parent compound and identification of major degradants. Where organic content of the soil is deficient, organic amendments should be supplied. The test should be conducted at a constant temperature between 18 and 30 °C. No specific requirements concerning light regime are provided. Data collection should be conducted until patterns of decline of the test substance can be detected, or for one year.

*Derivation of persistence endpoints:* No recommendations on the derivation of persistence endpoints are given.

#### 3.2.1.7 Other regulatory test systems

The Society of Environmental Toxicology and Chemistry (SETAC) published a guidance document containing “procedures for assessing the environmental fata and ecotoxicity of pesticides”, which were intended to be interim and to be replaced by the OECD guidelines (SETAC 1995).

Further, a Dutch guideline exists for the submission of applications for registration of pesticides, which describes a water-sediment test system. The guideline is written in Dutch and no translation is available (Cuwvo 1999).

#### 3.2.1.8 Modifications of regulatory test systems

Several studies have been published that present modified versions of OECD 308 and 309 test systems. These modifications were made to investigate the effect that changing some of the expected major influencing factors had on test outcomes. The most relevant studies that explored some of the effects of such modifications while staying close to the regulatory test systems in all other aspects are discussed in Chapter 3.3 for each factor in the section “Implementation in test system”.

### 3.2.2 Complementary experimental systems described in the scientific literature

#### 3.2.2.1 Mass balances in lakes and rivers

*References:* (Tixier et al. 2002, Tixier et al. 2003, Fono et al. 2006, Huntscha et al. 2008)

*Field of application:* Chemical degradation studies in real rivers and lake systems

*Objective:* Quantification of transformation half-lives in real rivers and lake systems

*Test system setup:* For the quantification of persistence in the field, a dedicated measurement campaign is usually carried out. Such measurement campaigns aim at quantifying all substance flows in the system under investigation. This typically requires the measurement of substance concentrations in the in- and outflow of the system and in the system itself (temporally and spatially resolved, if needed), as well as quantification of water flows (i.e., discharge) during the investigation period. Over

short river stretches and in the absence of sources along the river stretch investigated, it might be sufficient to measure concentrations at the beginning and end of the river stretch. The same is true for stratified lakes with little spatial heterogeneity and no direct sources into the lake (e.g., wastewater treatment plant effluents). In larger, stratified lakes, depth-dependent concentration profiles need to be measured along with temperature profiles to determine the stratification profile.

*Derivation of persistence endpoints:* More or less complex mass balance models are needed to derive transformation half-lives from the measured concentration and discharge data. In principle, if all substance flows into and out of the system have been sufficiently quantified and any fate processes other than transformation (e.g., sediment, volatilization) are sufficiently understood to be fully quantified, the rate constant of overall substance transformation can be estimated. Typically, the half-lives derived in this way, would, however, still lump together different transformation processes (i.e., transformation in both the water column and the sediment, phototransformation and biotransformation).

#### 3.2.2.2 Tracer and benchmarking approaches in lakes and rivers

*References:* (Johansson et al. 2001, Radke et al. 2010, Zou et al. 2014, Zou et al. 2015a, Zou et al. 2015b)

*Field of application:* Chemical degradation studies in real rivers and lake systems

*Objective:* Quantification of transformation half-lives in real rivers and lake systems

*Test system setup:* Quantification and modeling of complete mass balances can be very cumbersome, especially if there is much spatial and temporal variation in the system under investigation. Benchmarking approaches try to circumvent this by including benchmark chemicals into the selection of substances to be monitored. Suitable benchmark chemicals are substances whose environmental fate is well understood and that differ from the chemicals under investigation in only the property of interest, e.g., speed of transformation in the compartment of concern (i.e., persistence). The most straightforward application of this approach would be the use of conservative benchmark chemicals (i.e., non-sorbing, non-degradable substances), whose concentration profiles can be used to derive the extent of dilution without the need to quantify water flows (Radke et al. 2010, Kunkel et al. 2011, Zou et al. 2014, Zou et al. 2015a, Zou et al. 2015b). Fluorescence/dye tracer substances (e.g., uranine and rhodamine) can additionally be used to follow a parcel of water down a river such that downstream locations can be sampled upon passage of tracer substances to determine the fate of selected test chemicals contained in the same water parcel (Kunkel et al. 2011, Schwientek et al. 2016).

*Derivation of persistence endpoints:* If conservative benchmark chemicals are present in or have been added to in the system under investigation, persistence can be quantified based on the concentration ratio of a given test chemical and the benchmark chemical in (i) the medium that is the major vector of chemical input to the system (e.g., inflowing water or an emission source) and (ii) the water flowing out of the system. For the approach to work, it is essential that these ratios do not vary significantly over time. Derivation of the half-life of the test chemical then additionally requires either knowledge of the hydraulic retention time in the system (e.g., in the case of a lake) or the travel times (e.g., in the case of a river system). The latter can, for instance, be derived from experiments with tracer substances or from hydraulic modeling.

#### 3.2.2.3 Limnocorrals

*References:* (Solomon et al. 1985, Liber et al. 1997)

*Field of application:* Chemical degradation studies in real lake waters

*Objective:* Studying chemical fate in enclosures separated-off from the natural water body

*Test system setup:* Limnocorrals are transparent containments made of various inert materials (e.g., fiberglass, nylon reinforced polyvinyl chloride plastics) that extend from above the water's surface and

are fitted over lake sediment (Schrader et al. 2000). The cylinders settled into the sediment form a watertight seal, isolating the internal environment (water and sediment) from the pond. Within this walled-off column of lake water, treatments and manipulations can occur. Limnocorrals allow some control and replication yet being *in situ* very closely mimic ambient lake conditions such as photoperiod, temperature, light intensity, and native aquatic organisms.

*Derivation of persistence endpoints:* Half-lives so far were mostly determined from dissipation of the parent compound from the water column. Separate analysis of sediment content and samples of the containment wall controlled for losses other than transformation.

#### 3.2.2.4 Flume experiments

*References:* (Kunkel et al. 2008, Li et al. 2015)

*Field of application:* Chemical degradation studies in flowing waters

*Objective:* Studying chemical degradation under conditions simulating natural flowing waters

*Test system setup:* Flume experiments are conducted in specifically designed bench-scale annular flumes containing water and sediment. Flow velocity in the flumes is regulated with a propeller to allow for a continuous transport of water and, optionally, sediment. Chemicals are spiked into the surface water of the flume and water samples are taken over time to determine the decrease in concentration. To avoid photodegradation, flume experiments are carried out in the dark. To study the influence of flow velocity and sediment surface structure, experiments have been run at different flow velocities inducing slow and fast exchange of surface water with sediment (Kunkel et al. 2008), and with two difference surface structures (i.e., experiment FLAT with even sediment surface and experiment RIPPLE with artificial sediment ripples (Li et al. 2015)).

*Derivation of persistence endpoints:* In the initial phase of the experiments, decreasing concentrations in surface water are partially due to mixing with sediment porewater. Therefore, before calculation of half-lives, measured concentrations of test chemicals in the water column were corrected for dilution, using fluconazole, a conservative benchmark chemical, as reference compound. Half-lives were then determined by fitting first-order kinetics to the corrected surface water concentrations.

## 3.3 Factors influencing persistence in natural surface waters

The review of existing experimental approaches to quantify persistence in natural surface waters led to the identification of seven main factors that are expected to influence measured persistence at the water-sediment interface and that may vary throughout natural surface water systems. In the context of this project, it is important to note that the different regulatory test systems described in Chapter 3.2.1 vary with respect to how tightly these different factors are controlled and/or at what values they are kept. Table 8 gives an overview over the control of the different factors in regulatory test systems. In Chapters 3.3.1 to 3.3.7, the seven factors are discussed from a theoretical perspective as to how and to what extent they can potentially influence persistence at the water-sediment interface. Additionally, for each factor, parameters are listed that are known or likely to be available for at least some surface water bodies in Germany, and that could potentially serve as (proxy) variables to predict the spatial distribution and extent of that influencing factor across German surface waters.

Table 8: Varying factors in regulated test systems

| Factor | OECD 308 | OECD 309 / OPPTS 835.3190 | OPPTS 835.3180 | Canadian Guideline | OECD 53 | American guideline | |
|---|---|---|---|---|---|---|---|
| **Lighting** | Experiments carried out in the dark | Test should take place in the dark or under diffuse light | Physical parameters such as lighting, mixing, and temperature may be controlled to simulate the environmental conditions of the site from which water and sediment samples are collected. | Aerobic incubations should be carried out under standard lighting regime. If chemical is photolabile, the test should be carried out in the dark | Micro- and mesocosms are studied outdoors, under natural light conditions | No specifications | Physical / Chemical factors |
| **Water to sediment ratio** | Water-sediment volume ratio between 3:1 and 4:1 | Concentration of suspended sediments should be between 0.01 g/L and 1 g/L ("suspended sediment test") | | Water-sediment ratio not fixed | Depth of sediment should be >5 cm. | Water-sediment ratio not fixed | |
| **Temperature / Season** | 20±2 °C, where appropriate an additional lower temperature may be considered | Test should be carried out at 20-25±2°C | | Test should be carried out at two temperatures, one in the higher (20-30 °C) and one in the lower (3-8 °C) range | Micro- and mesocosms are studied outdoors, with natural variations in temperature | Soil-sediment should be maintained at any constant temperature between 18 and 30°C. | |
| **Disturbance of water and sediment layers** | Sediment to be disturbed as little as possible | Agitation by means of continuous shaking or stirring must be provided to maintain particles and microorganisms in suspension | | Determination of degradation either under static conditions or aerated by shaking | Disturbance to the system should be minimized | No specifications | |
| **Spiking of test chemical** | One-time application as aqueous solution into the water phase of the test system. For crop protection products, the spike concentration should be based on the maximum dosage on the label. | No statement on frequency of application of test chemical. Chemical should be applied as stock solution. At least two different concentrations of test substance differing by a factor of 5 to 10 should be used. Both of the | Test compounds are introduced into the system either as a single dose or by continuous dosing over the duration of testing. The test compound concentration shall approximate the expected ambient environmental concentration. | Pesticides should be applied at one or two dosage rates and be added to the water phase as a filter-sterilized aqueous solution. If two rates are done, a 10-fold difference is standard. Pesticide concentrations tested should be based | Direct application or "mimicking the route of entry" (e.g., drift and direct overspray can be simulated by doing a spray application). | No specifications | |

| Factor | OECD 308 | OECD 309 / OPPTS 835.3190 | OPPTS 835.3180 | Canadian Guideline | OECD 53 | American guideline | |
|---|---|---|---|---|---|---|---|
| | In all other cases, the concentration to be used should be based on predictions from environmental emissions. | selected concentrations should be less than 100 µg/L and preferably in the range of <1-10 µg/L. | | on the maximum label recommended rate or a concentration expected to occur in water runoff or as a result of spray drift (generally < 1 µg/mL). | The selected concentrations should generally be based on those expected to cause effects in the system. | | |
| **Sediment composition** | Two sediments should be used that differ in organic carbon content and texture. For the anaerobic studies, the sediments should be sampled from the anaerobic zones of surface water bodies. | Sediment should come from the same site as that from which the water sample was taken. The sediment may either be characterized by a high organic carbon content (2.5-7.5%) and a fine texture or by a low organic carbon content (0.5-2.5%) and a coarse texture. | Water and sediment samples should be collected from the same site. | The guideline does not provide regulations concerning the origin or structure of the sediment. It only states that the report should contain information on textural class, particle size distribution and % of organic carbon of the sediment as well as the geographical location. | Sediments can be collected from supply ponds or natural systems. Alternatively, soil can be used that has been conditioned sufficiently to have aquatic sediment-like properties. | Preferred substrate for this laboratory study is sediment covered with water, but the use of a flooded soil is also considered adequate. | |
| **Additional biology** | No additional biology is included in the existing experimental setup | No additional biology is included in the existing experimental setup | No additional biology is included in the existing experimental setup | No additional biology is included in the existing experimental setup | Algae, macrophytes, invertebrates and fish are included into the test system | No additional biology is included in the existing experimental setup | **Biological factors** |

### 3.3.1 Lighting

*Theory:*

Light might influence degradation in water-sediment systems in at least two ways. First, light might induce phototransformation of chemicals, a process that can further be subdivided into direct and indirect phototransformation. In direct phototransformation, the chemical directly absorbs light and as a consequence undergoes transformation. In indirect phototransformation, the light is absorbed by naturally occurring components of the water matrix (e.g., dissolved organic matter, nitrate, nitrite), which may lead to the formation of reactive species that initialize the transformation of chemicals (Schwarzenbach et al. 2005). Second, light is needed for the growth of photoautotrophic organisms such as algae or macrophytes. These may directly or indirectly, through their associated microbiomes and photosynthesis-related pH changes, contribute to the transformation of chemicals.

*Environmental relevance:*

Phototransformation can significantly influence the degradation rate of chemicals in aqueous systems. OECD 309 tests carried out for a specific compound in the dark and under irradiated conditions resulted in significantly different half-life times. For experiments in the dark, an average $DegT_{50,ts}$ of 122.8 hours was calculated whereas under the influence of light a $DegT_{50,ts}$ of 1.4 hours was found (UBA 2016). Similarly, flume experiments with diclofenac that were carried out in the dark resulted in half-lives of 3.2 - 8.5 d (Kunkel et al. 2008), whereas experiments with fortified lake water exposed to sunlight showed rapid phototransformation with (pseudo) first-order kinetics and a half-life of less than 1 h (Buser et al. 1998). Rua-Gomez et al. (Rua-Gomez et al. 2013) investigated phototransformation and biotic transformation of three pharmaceuticals and two of their major transformation products in surface waters. They found that, for the sunlit, top layer of surface water, degradation by indirect phototransformation was much faster (i.e., half-lives of 1-5 d) than biotransformation (i.e., half-lives of around 100 d).

In general, because the electronic absorption spectrum of most chemicals shows only little overlap with the spectrum of terrestrial sunlight (mainly in the Vis and UV-A range), only a few chemicals are affected by direct phototransformation (e.g., trifluralin, a dinitroaniline derivative, which absorbs sunlight even in the visible spectral region). By contrast, indirect phototransformation processes are more likely, because various photochemically active light absorbers are present in surface waters. Thorough overviews of the existing literature on the phototransformation of pharmaceuticals and pesticides in surface water bodies are given in (Boreen et al. Boreen2003 2003) and (Burrows et al. 2002), respectively. With respect to the importance of phototransformation on overall degradation in surface water bodies, one additionally needs to consider the fact that the effective depth of the photoactive layer is typically in the range of maximally 0.5-1 m. This value may be lower in waters with high concentrations of organic material and/or phytoplankton that absorb light. Thus, the depth of surface water bodies and the composition of the water matrix will have a large impact on the contribution of phototransformation to overall degradation (Vione et al. 2010).

The potential contribution of photoautrophic organisms to chemical degradation is discussed in Chapter 3.3.7.

*Implementation in test systems:*

The OECD 308 test system is the only system that strictly excludes any light. All other regulatory test systems listed in Table 8 allow for the presence of some diffuse light source or even natural light conditions.

*Potentially available information on relevant parameters in German surface waters:*

- River height

- DOC, nitrate, nitrite
- Turbidity, Chlorophyll A

### 3.3.2 Water-sediment ratio

*Theory:*

The water-sediment ratio might influence degradation in water-sediment systems in several ways. First, the sediment solid phase reduces the availability of chemicals for transformation as a function of their solid-water distribution ratio $K_d$. Second, part of the sediment organic matter (SOM) constitutes of active microbial biomass (Honti et al. 2016), which may biotransform the chemicals. Typically, for more polar compounds ($K_d < 10^4$ L/kg), the biomass effect prevails, causing a certain proportionality between the amount of sediment and degradation rates. Only for very hydrophobic compounds and in fully mixed systems (e.g., OECD 309), these opposing trends as a function of the amount of sediment may compensate each other and thus effectively cancel out the influence of the water-sediment ratio. In layered systems (e.g., OECD 308), the water-sediment ratio translates into a specific ratio of the column heights of the water and sediment compartments, respectively. If it is assumed that degradation is mostly by biotransformation and biotransformation mostly takes place in the top oxygenated layer of the sediment, an increased water-sediment ratio will significantly decrease overall degradation in the system because the chemical is redistributed into the larger water column and hence less available for biotransformation (Honti et al. 2015). Finally, it has been argued that a low water-sediment ratio may lead to unrealistically large formation of non-extractable residues because of excessive distribution of the chemical into the sediment (ECETOC 2010).

*Environmental relevance:*

Based on the above considerations, it can be assumed that the depth of the water column will have a very significant impact on how much biotransformation is observed overall in a surface water body. To the best of our knowledge, this has, however, not been experimentally confirmed in real surface water bodies.

*Implementation in test systems:*

Water-sediment ratios are recommended in both the OECD 308 guideline (i.e., recommended water-sediment volume ratio of 3:1 to 4:1) and the OECD 309 guideline (i.e., 0.01 to 1 g/L). In the OECD 308 guideline, the depth of the sediment layer is further specified at 2.5±0.5 cm. For the other regulatory test systems listed in Table 1, the water-sediment ratio is not explicitly specified. Recently, Shrestha et al. (Shrestha et al. 2016)have tested a modified version of OECD 308 with a 10:1 (w:w) water-sediment ratio. They also tested a modified version of OECD 309 containing a 10-fold higher amount of suspended sediment (i.e., 10 g/L) than the upper limit recommended by the guideline (i.e., 1 g/L). For the fully mixed systems, degradation in the modified OECD 309 system was generally higher than in the standard system, thus confirming the accelerating effect the higher amount of sediment had on degradation. Similar results had been obtained for increased suspended sediment concentrations in Rhine water previously (Wanner et al. 1989). For the OECD 308-type systems, the increased water-sediment ratio in the modified system was accompanied by an increased oxygenation of the sediment layer, making it difficult to interpret the results with respect to the separate influence of the water-sediment ratio.

Natural surface waters used in the original setup of the OECD 308 test to some extent always contain biologically active, suspended sediments that contribute to some extent to biotransformation. To explicitly limit biotransformation to the sediment compartment only, OECD 308 tests with synthetic river water have been carried out instead (Radke et al. 2011, Li et al. 2014, Radke et al. 2014) The synthetic

river water was added to the sampled sediment in glass bottles and they were equilibrated for one week before addition of the pharmaceuticals.

*Potentially available information on relevant parameters in German surface waters:*

- River height

### 3.3.3 Sediment composition and texture

*Theory:*

The sediment composition (i.e., particularly the fraction of SOM) and the texture may influence degradation in several ways. First, as described in Chapter 3.3.2, the amount of SOM influences bioavailability and the amount of active biomass. Sediment texture affects the exchange between the water column and the sediment, with coarser substrates typically fostering increased exchange fluxes (i.e., increased availability of oxygen, nutrients and chemicals to be degraded) and hence most likely increased biotransformation. Texture further affects the growth of biofilms through the provision of nutrients and support for growth (Battin et al. 2016).

*Environmental relevance:*

SOM may vary widely between sediments (from <1% up to 10%). Sediment textures to some extent correlate with SOM, i.e., fine sediments with high clay and silt content are typically related to high SOM, and coarse sediments with high sand content are typically related to low SOM. No studies could be found that specifically investigated the influence of sediment composition and texture on the biotransformation of chemicals.

*Implementation in test systems:*

In the OECD 308 guideline testing of two sediments, one with a high (2.5-7.5%) organic carbon content and a fine texture, and one with a low (0.5-2.5%) organic carbon content and a coarse texture, is required. In the guideline, “fine texture" is defined as a [clay + silt] content of >50% and "coarse texture" is defined as a [clay + silt] content of <50%. The OECD 309 guideline states that either a low or high SOM sediment could be used.

*Potentially available information on relevant parameters in German surface waters:*

- River bed composition and textural class

### 3.3.4 Disturbance of water and sediment layer

*Theory:*

Under different environmental or experimental conditions, the contact between the chemicals in the water phase and different parts of the sediment solids may vary a lot. Under stagnant conditions, a system of different layers forms including a water column with very little suspended sediment, an intermediate mixed layer of highly porous sediment material all the way to more compacted layers at higher depths. If the mixed layer contains active biomass, the layering will also result in a redox gradient from aerobic conditions in the top most mm of the mixed layer, to anoxic and finally anaerobic conditions in deeper layers. The depth profile of redox zonation in layered systems will depend on various factors, including sediment bed surface morphology, flow velocity in the water column, dissolved and particulate organic carbon concentrations etc. In highly disturbed systems (e.g., rivers with highly turbulent water flows), sediment particles may be resuspended in large amounts, leading to large concentrations of suspended biomass under fully aerobic conditions.

*Environmental relevance:*

With the exception of a few specific groups of chemicals (e.g., highly halogenated compounds, nitro compounds), biotransformation of most chemicals is more likely to proceed through oxidative transformation, i.e., to happen under aerobic conditions. Therefore, the absolute and relative amount of aerobic sediment can be expected to impact how much biotransformation is observed in a water-sediment system. This is also demonstrated by comparing the results of OECD 308 studies carried out under aerobic and anaerobic incubation conditions. Radke and Maier (Radke et al. 2014) showed that for 6 out of 8 chemicals, degradation half-lives clearly decreased when conditions where switched from anaerobic to aerobic. Thus, overall, more (aerobic) biotransformation would be expected in more disturbed systems, i.e., systems with a rugged sediment bed surface and high flow velocity. The depth of the oxic zone can also vary laterally across large riverbeds. For instance (Fischer et al. Fischer2005 2005) reported a very deep oxic zone in the deep channel in the middle of the river Elbe versus much thinner oxic zones in the more stagnant nearshore zones. Correspondingly the depth-integrated microbial productivity was up to five times higher in the central channel compared to the nearshore habitats. Li et al. (Li et al. 2015) have compared biotransformation of 19 pharmaceuticals in flume experiments with flat and rippled sediment bed morphologies. They did not observe a significant effect on dissipation half-lives. However, they neither observed faster equilibration times in the rippled sediment compared to the flat sediment, suggesting that the expected increased exchange of water between sediment and surface water with the rippled surface had not actually taken place (most likely due to a non-ideal distance between ripples in the experimental setup).

*Implementation in test systems:*

OECD 308 requires a stagnant water phase with as little disturbance as possible. Shrestha et al. (Shrestha et al. 2016)have tested a modified version of OECD 308 in which the water phase was slowly stirred from above. They showed that this led to > 2 mm of the sediment being fully oxygenated, compared to only about 1 mm in the standard OECD 308 setup. This also went hand-in-hand with an increased degradation in the modified test system, despite the higher water-sediment ratio. These results show that disturbance can have a large impact on degradation, and that, therefore, the conditions in OECD 308 are most likely not suitable to simulate conditions in flowing water (e.g., rivers) or the open sea as acknowledged in the respective guideline under point 6 (OECD 2002).

*Potentially available information on relevant parameters in German surface waters:*

- Discharge
- Flow velocities

### 3.3.5 Temperature/seasonality

*Theory:*

Two major factors changing with season and/or geographical region are the outside temperature and the intensity and duration of sunlight periods. For a discussion of the influence of sunlight on degradation in water-sediment systems, see 1.1.1. Temperature might influence degradation in water-sediment systems in two ways. First, the Arrhenius law states that chemical degradation is temperature-dependent and that the magnitude of change with temperature depends on the activation energy ($E_a$) (of the slowest step of the sequence) of the transformation reaction(s) (Schwarzenbach et al. 2005). This certainly applies to all chemical reactions (i.e., hydrolysis, photolysis), but theoretically also applies to enzyme-catalyzed biotransformation reactions. However, in the case of microbial biotransformation reactions, a second influence needs to be considered. Temperature and/or season can also shape the composition of the microbial community (Wells et al. 2011), which, in turn, might strongly

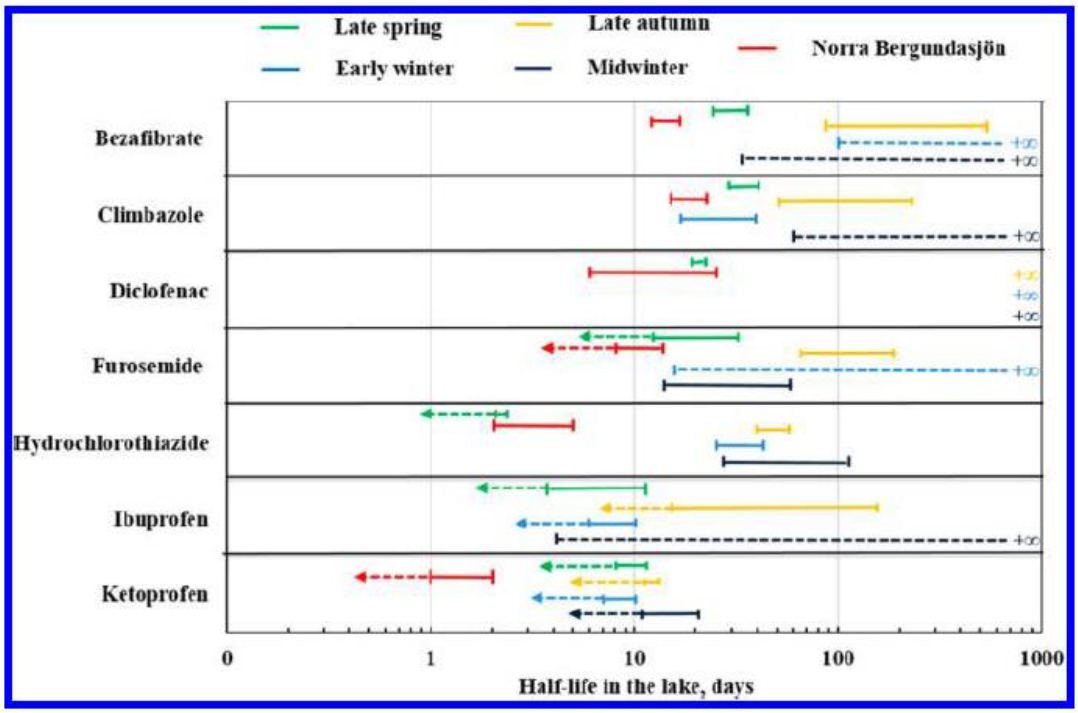
impact the capacity of the community for degrading specific chemicals. Finally, temperature and seasons will also affect discharge, particularly in rivers fed by water from snowmelt. In those rivers, higher discharge and disturbance are typically observed during summer. The higher discharge will lower the chemicals' residence time in the river system and hence lead to less overall removal.

*Environmental relevance:*

Theoretically disentangling the Arrhenius effect from more complex effects of changing community composition is not practical. For pesticides, a more pragmatic approach of surveying available soil degradation studies to derive $E_a$ values was chosen. The responsible panel recommended that a median $E_a$ value of 65.4 kJ mol$^{-1}$ was used for extrapolating pesticide degradation in soil between different temperatures (Residues 2008). This corresponds to a factor of 2.58 difference for a 10°C-change in temperature ($Q10$). However, they also acknowledged the fact that compound-specific differences in $E_a$ exist and that it is hence not fully correct to assume that there is one median $E_a$ value for all pesticides. Since microbial biotransformation is responsible for the degradation of most pesticides in soil, a similar factor can, on average, be expected to apply in water-sediment systems.

Zou et al. (Zou et al. 2015b) used a chemical benchmarking approach to investigate the temporal variation of the persistence of chemical contaminants in a Swedish lake. For the 7 quantified chemicals, transformation half-lives varied over 1-3 orders of magnitude between different seasons (Figure 18). For 5 of the chemicals, the measured half-lives were lower in late spring than in late autumn/early winter. While for the two chemicals with the most extreme differences (i.e., diclofenac, hydrochlorothiazide) photochemical transformation most likely played a major role, the observed difference for the other compounds, which was typically around a factor of 3-10 between seasons, can likely be at least partially attributed to temperature-dependent biotransformation.

Figure 18: Estimated half-lives of seven pharmaceuticals during 4 study time periods in Boren and during late spring in another Swedish lake (Norra Bergundasjön) (Zou et al. 2015b)




*Implementation in test systems:*

The Canadian guideline on persistence of pesticides is the only guideline that specifically requests consideration of varying temperature and seasonality by requiring the tests to be conduct at two different temperatures (high and low). The other test guidelines mostly neglect this aspect by recommending only one test temperature. In the OECD 308 guideline, conducting an additional experiment at a lower temperature is mentioned as optional where "appropriate".

*Potentially available information on relevant parameters in German surface waters:*

- River temperatures

### 3.3.6 Spiking of test chemicals

*Theory and environmental relevance:*

The amount of test chemical spiked and the spiking frequency can influence degradation kinetics in the water-sediment system. According to Michaelis-Menten enzyme kinetics, the extent and order of chemical degradation kinetics changes with changing substrate concentrations (Battersby 1990, Schwarzenbach et al. 2005). At low substrate concentrations typically encountered for chemicals in aquatic environments, kinetics are generally second order relative to the substrate concentration and the active microbial biomass. Since, at low substrate concentrations growth of biomass is unlikely, this often simplifies to a pseudo-first order rate law that is only proportional to the substrate concentrations. If, however, unrealistically high substrate-to-biomass concentrations are spiked into the test systems, kinetics may change to zero order kinetics and/or biomass growth may be observed, leading to Monod population growth-type kinetics (Battersby 1990, Schwarzenbach et al. 2005) . Similarly, if the compound is spiked repeatedly at elevated concentrations, an increase in biotransformation rates over time or after an initial lag phase might be observed due to adaptation of the population.

*Implementation in test systems:*

While the OECD 308 explicitly recommends a one-time spiking at the beginning of the experiment, other guidelines (e.g., OECD 309) do not provide specific statements about the spiking frequency of test substances or recommend multiple spiking or continuous dosing. Most guidelines with the exception of OECD 53, which is targeted towards effect testing, recommend applied concentrations to either represent environmentally relevant concentrations or to be consistent with the maximal recommended dosing rate.

### 3.3.7 Additional biology

*Theory:*

Besides the microbial biomass, all other living organisms present in aquatic systems may also be able to biotransform a given chemical to some extent. However, mass balance-wise, it is particularly the potential contributions of macrophytes (i.e., plant biomass) and photosynthetic microorganisms, besides the hetero- and chemoautotrophic microbes, that should be of interest. These communities may not only directly metabolize several chemicals, but they are also capable of enhancing the diversity, activity and amount of biomass of the microbial community (Thomas et al. 2011).

*Environmental relevance:*

Depending on nutrient conditions, light penetration and substrate morphology, surface water sediments are covered to some extent by stream biofilms, which are comprised of multiple species of algae, heterotrophic and chemoautotrophic bacteria, cyanobacteria, and other microorganisms within a self-developed polysaccharide matrix (Lock et al. 1984, Battin et al. 2016). Those so-called periphyton communities play a key role in the carbon and nutrient dynamics of stream ecosystems, and hence also

a key role in the degradation of chemical contaminants. It is further well established that the dissolved organic carbon (DOC) released by the algae supports growth of the heterotrophic biomass, resulting in a tight coupling between algae and bacteria, especially under low available nutrient conditions in benthic biofilms (Lyon et al. 2009). It seems intuitive, but has also been demonstrated experimentally, that exclusion or inclusion of algae into biodegradation testing would yield different results. In terms of experimental evidence, Jasper et al. (Jasper et al. 2014,), for instance, found an increased transformation of two pharmaceuticals in wetland microcosms when algal photosynthesis was supported by irradiation with visible light. Thomas and Hand (Thomas et al. 2012) tested different fractions of a benthic periphyton for their capacity to degrade a specific fungicide, and found both the algal fraction as well as the isolated macrophytes to be more active than the heterotrophic biomass. With respect to macrophytes, (LeFevre et al. 2015) also reported on the rapid phytotransformation of benzotriazoles by the model plant Arabidopsis. They mainly attributed this to different conjugation reactions they observed, which is a type of metabolic pathway that is much more strongly associated with eukaryotes than with prokaryotes (i.e., bacteria). Complete exclusion of photoautotrophic organisms from water-sediment degradation studies thus not only potentially excludes important growth factors for the microbial biomass, but also certain biotransformation pathways that may lead to rapid transformation of certain chemical contaminants.

*Implementation in test systems:*

Algae and macrophytes are only considered in the outdoor micro- and mesocosm test systems included in the OECD Nr. 53 guideline. All other guidelines advocate for tests to be carried out in the dark, thus effectively eliminating the possibility for photoautotrophic organisms to be present. Hand and co-workers (Hand et al. 2010, Thomas et al. 2011) tested modified versions of OECD 308 where they added algae and macrophytes to a water-sediment system under a regime of fluorescent light (which, along with the use of borosilicate glass, excluded ultraviolet wavelengths and, hence, minimized degradation by photolysis) on a 16:8 h light:dark cycle. They found that including macrophytes or algae resulted in significantly increased rates of degradation of six out of altogether seven pesticides tested with differences in degradation half-lives as high as one order of magnitude. (Hand et al. 2010, Thomas et al. 2011).

*Potentially available information on relevant parameters in German surface waters:*

- Coverage and thickness of benthic biofilm
- Light penetration

## 3.4 Variability in substance half-lives across laboratory systems

Even though several studies were conducted on the degradation of chemicals in water-sediment systems, the number of studies testing the same chemical is limited. In Appendix 1, degradation half-lives for pharmaceuticals with half-lives in water-sediment systems reported in the peer-reviewed literature between 2005-2015 have been compiled.

Table 9 shows an excerpt of $DegT_{50,ts}$ values from Appendix 1. It includes data measured in different laboratory-based test systems for diclofenac, ibuprofen, naproxen and propranolol, which were the only four compounds with reported results from three or more laboratory studies. The studies conducted by (Araujo et al. 2011) and (Araujo et al. 2014) were performed with surface water only, without inclusion of sediment into the test system. Although there is some overlap in half-life ranges with those measured in the presence of sediment, the upper limit of half-lives measured in water only systems is typically higher. These limited data thus seem to confirm that the presence of sediment biomass does increase the rates of biotransformation (see Chapter 3.3.2 on water-sediment ratio). Outcomes from different laboratory test systems with sediment (including standard OECD 308) show comparably more

agreement ($DegT_{50,ts}$ values well within a factor of four or less, with the exception of the upper limit value for ibuprofen in (Radke et al. 2011)). When comparing only across OECD 308 studies, experimental results are even more similar. For 17 pesticides (UBA 2016), the ratios between $DegT_{50,ts}$ values for two different sediments and the same compound varied between 1.1 and 2.5, with a median value of 1.5. For 74 pharmaceuticals for which studies with multiple sediments were available (UBA 2016), only four compounds exhibited ratios of $DegT_{50,ts}$ values between two sediments of >10. The median ratio between $DegT_{50,ts}$ values for two different sediments and the same compound across all pharmaceuticals was 2.0.

**Conclusion**

These limited data thus suggest that in the presence of sediment, in the dark and within reasonable variations in experimental setups, differences of less than an order of magnitude in $DegT_{50,ts}$ values for a given compound can be expected.

Table 9: $DegT_{50,ts}$ values from different laboratory test systems for diclofenac, ibuprofen, naproxen and propranolol.

| | (Kunkel et al. 2008) | (Li et al. 2015) | (Löffler et al. 2005) | (Radke et al. 2009) | (Radke 2011) | (Caracciolo et al. 2012) | (Araujo et al. 2011) | (Araujo et al. 2014) |
|---|---|---|---|---|---|---|---|---|
| **System type** | **Flume experiment** | **Flume experiment** | **OECD 308** | **Sediment with artificial river water** | **Sediment with artificial river water** | **Microcosm** | **Surface water only** | **Surface lake water** |
| **diclofenac** | 3.2 - 8.5 | | | 11.3 | | | | 2.14 ± 0.2 - 60 ± 9.4 |
| **ibuprofen** | 1.2 - 2.5 | 1.8 | < 6 | | 2.4 - 45 | | | 17.8 ±2.1 - 247.7 ± 30.1 |
| **naproxen** | 5.4 - 6.9 | | | | | <22 | 10.2 ± 0.5 - 14.6 ± 1.3 | |
| **propranolol** | | 5.6 - 7.1 | | 11.1 | 3 - 14 | | | |

## 3.5 Environmental representativeness of OECD 308 & 309 test systems

To address the question about the environmental representativeness of OECD 308 and 309 test systems, one needs to understand in what form the above-discussed influencing factors are actually present in natural river systems in Germany. To illustrate this point, Table 10 provides an overview of four highly prevalent, yet distinctly different river types in Germany and some of their major characteristics that are either directly or indirectly linked to those influencing factors and are hence thought to affect

biotransformation of chemicals. The classification of river systems was adapted from (Pottgiesser et al. 2008).

Table 10: Four typical German river types and their major characteristics that are potentially relevant for biotransformation extracted from (Pottgiesser et al. 2008)

| Category | Sand-bottomed lowland rivers | Carbonic low mountain streams with enriched fine material content | Gravel dominated streams | Streams of the alpine foreland |
|---|---|---|---|---|
| River type according to (Pottgiesser et al. 2008) | 14 | 6 | 10 | 2 |
| Morphology | Highly meandering in flat valley | Meandering | Meandering | Meandering |
| Sediment composition | Sand, gravel | Clay, loess, silt and fine sands | Gravel | Gravel |
| Size of catchment area | 10 - 100 $km^2$ catchment area | 10 - 100 $km^2$ catchment area | > 10.000 $km^2$ catchment area | 10 - 1000$km^2$ catchment area |
| Valley floor gradient | 2-7 ‰ | 4-30 ‰ | 0.2 -2 ‰ | > 0.5 ‰ |
| Macro-zoobenthos | Little colonized fine-grained sediment | Increased proportion of colonized fine-grained sediment | Little colonized fine-grained sediment | Dominated by rheophile stone colonisers |
| Phytoplankton | No plankton | No plankton | Little plankton | No plankton |
| Examples | Rotbach | Fischbach, Tiefenbach | Elbe | Sempt, Nöbach |

From Table 10, it can clearly be seen that a given river system can have different characteristics that are most suitably represented by features of either an OECD 308 or 309 test, and that while one characteristic might be best represented by one of the test systems, another characteristic of the same river might be more suitably represented by the other test system. A few aspects are exemplarily discussed in the following:

- The stagnant conditions in the OECD 308 test system are typically not suitable to represent conditions in flowing water (OECD 2002). Nevertheless, it might still be fairly representative of a type 10 river because of its typically low floor gradient and therefore low flow velocity. For most other river systems, e.g., for a type 6 stream that is characterized by a higher floor gradient or for less meandering rivers with higher flow velocity, an OECD 309 test system with some suspended sediment is most likely more representative.

- The fact that OECD 308 studies are always run in the dark might make them more representative of river systems that are deep and/or inhabited by plankton and therefore are less transparent for light (e.g., type 10 river).
- A sediment with a fine substrate, e.g., as present in type 6 rivers, typically has a higher soil organic matter (SOM) content and a higher biomass density. This typically also leads to a steeper redox gradient in the sediment. Both conditions are better represented by an OECD 308 test system compared to the OECD 309 test system.
- While rivers with small catchment areas typically are rather shallow and might therefore be reasonably well represented by the water-to-sediment ratio of 3:1-4:1 used in an OECD 308 test system, larger streams are typically deeper and might be better represented by the higher water-to-sediment ratio of the OECD 309 test system.

Thus, while the degree of disturbance and the lack of light might suggest that an OECD 308 test system is most suitable to represent a type 10 river, the gravel-dominated and thus well aerated sediment bed and the considerable depth of those types of rivers are characteristics that would most likely be better represented by an OECD 309 system.

**Conclusion**

The considerations presented underline that there is no single test system that best represents the conditions in German river systems in general, nor might there even be a single test system that is most representative of a given type of river system in all aspects relevant to biotransformation.

Given the above-said, an alternative approach in a regulatory context is to choose a test system that errs on the safe side, i.e., that is sufficiently but not over conservative. Laboratory test systems are generally non-conservative with respect to temperature (20-25°C). As a consequence, there is now a new requirement within REACH to carry out the OECD 308 and 309 studies also at a lower temperature of 12°C (ECHA 2016a) Between OECD 308 and 309, it seems that the OECD 308 test system is more conservative in at least two aspects: First, it specifically requires the test to be carried out in the dark, and hence no phototransformation or growth of autotrophic organisms can take place. Second, the system should be disturbed as little as possible, leading to a very thin oxic layer only. In 309, diffuse lighting is allowed (i.e., many irradiated studies have been submitted (UBA 2016), and the system is highly disturbed (i.e., available biomass is most likely experiencing fully oxic conditions). However, there is also at least one aspect where OECD 308 is actually clearly less conservative than 309, and that is the water-sediment ratio, which is much lower in 308. Thus, from a total mass balance perspective, more of the chemical is absorbed into the sediment phase and thus in direct proximity to where the active biomass is located. Interestingly, recent results by (Shrestha et al. 2016) actually suggest that the extent of degradation for two degradable compounds is quite comparable between the OECD 308 and 309 system, but for different reasons. Also, the same study showed that increased disturbance in a modified version of the 308 system and increased sediment concentrations in a modified version of the 309 system both significantly increased degradation.

**Conclusion**

Overall, these examples show that prediction of the quantitative influence of the different influencing factors on test system outcomes is currently not or only partially feasible. We therefore cannot predict which test system would ultimately yield more conservative results. Moreover, such relative outcomes between test systems would potentially be different for chemicals with strongly deviating properties (e.g., for strongly sorbing chemicals or for chemicals that are preferentially transformed under anaerobic conditions).

## 3.6 Discussion and recommendations

In conclusion, theoretical evaluation of the different test systems and their divergent properties with respect to factors that can be expected to influence biotransformation clearly demonstrates that variability between test system outcomes is unavoidable. It is therefore strongly recommended that in the regulatory process explicit and transparent strategies are needed to deal with this. Three different strategies to do so can be perceived.

A first strategy is to choose the appropriate test system on a substance-by-substance basis such that, based on substance properties and emission scenarios, it best represents the exposure compartment where most of the substance mass will reside. An example of this is the current recommendation in the REACH Information Requirements (ECHA 2016b) for strongly sorbing substances to prefer testing according to OECD 308 over OECD 309. This is because these types of substances are expected to mainly reside in the sediment compartment in natural aquatic systems, and this is consistent with what happens in the OECD 308 test system where the major part of their mass will be immediately transferred to the sediment. For these strongly sorbing substances, it can be shown that $DegT_{50,ts}$ is equal to $DegT_{50,sed}$ (see Chapter 2.1.2.3), and that the total system half-life thus is a good indicator of persistence in sediment. However, for substances with intermediate sorption behavior, it is more difficult to choose the test system and appropriate persistence indicator based on these principles because their mass distribution may shift significantly between the sediment and water column, depending on water column height and sediment properties.

A second strategy, therefore, is to extract more fundamental information on biotransformation that is less dependent on the actual test system geometry and water:sediment ratio from the outcomes of any test system. In this way, at least some of the causes for the observed variability between outcomes from different test systems can be eliminated. The suggestion to derive a biomass- and bioavailability-normalized biotransformation rate constant $k'_{bio}$ as suggested in Chapter 2.1.3is consistent with this strategy.

Finally, a third strategy would be to use the outcome of any test system that is either an OECD 308, an OECD 309 or a variant thereof for persistence assessment. In this case, the assessment would need to acknowledge, however, the uncertainty of at least a factor of 10 inherent in doing so. This could, for instance, be done by implementing a corresponding “safety factor”, either in the persistence criterion itself or by multiplying the test outcome prior to comparison to the persistence criterion. While this last strategy would most likely be easily implementable and consistent with other practice in PBT and risk assessment, it remains the least transparent. This is even more so since the rationale behind and the protection aims of the current persistence criteria remain unclear (Matthies et al. 2016).

# 4. Suitability of regulatory data to predict fate in rivers

In this chapter, the question is addressed how comparable the half-lives measured in regulatory tests are relative to degradation half-lives observed in actual surface water bodies. This question is addressed in two ways: First, literature-reported half-lives are compared for substance for which half-lives are reported from both laboratory studies and field studies (Chapter 4.1). Second, a field study in the Rhine river where the fate of diverse micropollutants in a parcel of water was followed down the Rhine is used as a case study. The measured concentrations in this parcel of water, as it was traveling down the Rhine, were used to estimate half-lives in the Rhine, which could then again be compared to the half-lives derived from OECD 308 data (Chapter 4.2).

## 4.1 Comparison of laboratory half-lives with field half-lives from the scientific literature

Table 11 shows an excerpt of literature-reported $DegT_{50,ts}$ values as compiled in Appendix 1. Here, ten compounds have been selected for which at least one measured half-life in both a laboratory test system and from a field study in a real river or lake system was available. Given the limited amount of available data, no statistically valid comparison of half-lives measured in laboratory test systems and those measured directly in the field is possible. One potentially relevant observation is the fact that the ratio between the half-lives observed in the field and those measured in laboratory test systems varies between 0.19 and 8.4 only, and does thus not exceed a factor of 10 for any compound. One important factor potentially influencing this comparison is the fact that some of the degradation in the real system might be due to photochemical transformation. We have therefore collated qualitative information from the scientific literature as to the photochemical degradation potential of the respective compounds. This information has been added to Figure 20. However, no clear trend towards lower field-to-laboratory half-life ratios in the case of compounds with known high photochemical degradation can be recognized.

Table 11: $DegT_{50,ts}$ values from laboratory test systems and field studies in real rivers or lakes for pharmaceuticals for which both types of values were available.

| | **Laboratory studies** | | | | | **Field studies** | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $DegT_{50,ts}$ (d) | (Kunkel et al., 2008) | (Li et al., 2015) | (Loeffler et al., 2005) | (Lam et al., 2004) | (Caracciolo et al., 2012) | (Radke et al., 2010) | (Kunkel et al., 2011) | (Zou et al., 2015b) | | |
| | Flume experiment | Flume experiment | OECD 308 | Microcosm | Microcosm | Tracer | Tracer | Chemical Benchmarking | Photochemical degradation potential (based on references in Appendix 2 | Ratio Field to Laboratory ratio (average Field / Average lab) |
| bezafibrate | 2.5 - 4.3 | | | | | | | 14 | - | 4.12 |
| diclofenac | 3.2 - 8.5 | | | | | | | 10 - 13 | High | 1.92 |
| gemfibrozil | 5.6 | | | | | | | 47 | Medium | 8.39 |
| ibuprofen | 1.2 - 2.5 | 1.8 | < 6 | | | | 10 ± 1.3 h | | Low | 0.22 |
| naproxen | 5.4 - 6.9 | | | | <22 | 3.6 ± 2.1 | | | High | 0.44 |
| clofibric acid | | 12-14 | | | | | 2.5 ± 0.5 | | Medium | 0.19 |
| furosemide | | 15 - 16 | | | | | | <10 | - | 0.32 |
| ketoprofen | | 2 -2.2 | | | | | | <2 | High | 0.48 |
| sulfamethoxazole | | 34 | | 19 ± 1.2 | | | | 26 | High | 1.02 |
| carbamazepine | | | 328 | | | | | 1200 – 1400 | Low | 3.96 |

## 4.2 Using mass balance modeling to derive persistence in the river Rhine

Here, the goal was to estimate persistence in a large river and to compare it to persistence estimated from OECD 308 simulation studies. For this purpose, concentration data measured in the Department of Environmental Chemistry at Eawag for several micropollutants in a parcel of water that has been followed down the Rhine and analyzed in several locations ("Rhine wave"(Ruff et al. 2015)) was compared with predictive modelling of the behavior of selected ones of those compounds in the Rhine wave.

## 4.2.1 Methods

### 4.2.1.1 Selection of substances

Based on the objectives of the study, the following selection criteria for substances were applied:

- OECD 308 data should be available for these substances
- The substances chosen should be pharmaceuticals because their spatial emission pattern can be predicted with much more accuracy than for pesticides
- The substances should have been measured in the Rhine wave

Four substances fulfilled all of these criteria. Because the OECD 308 data were confidential, these four substances were anonymized. We further included carbamazepine as a conservative benchmark chemical. Carbamazepine is widely reported to undergo only minor degradation and to be suitable for use as a chemical tracer (Clara et al. 2004, Fenz et al. 2005, Nakada et al. 2008, Bahlmann et al. 2009). Therefore, although we had no OECD 308 data for carbamazepine, we assumed it to represent the case of a highly conservative control substance. Last, we also included two substances that we suspected to be potentially highly degradable based on their concentration patterns along the Rhine (trimethoprim and sitagliptin). The latter three substances were not anonymized because no OECD 308 data were available for them anyway.

Finally, the following three groups of substances were defined and used in predictive modeling:

Group I: Conservative substance (carbamazepine)

Group II: Substances with OECD 308 data available (substance API6, substance API8, substance API9, substance API13)

Group III: Substances with assumed high degradation (trimethoprim, sitagliptin)

In Table 12, the total system and compartment-specific half-lives of Group II substances as calculated according to the methods outlined in Chapter 2.1.1 are summarized. The degradation rate constants used for modeling of Rhine concentrations were derived from these half-lives. The values given in Table 12 demonstrate that for all substances half-lives for the total system are significantly lower than for the water compartment. It is further shown that substances API8, API13 and API6 degrade rapidly in the oxic layer of the sediment, with half-lives <4 days. This is not the case for substance API8, which is indicated to be very persistent in the aquatic environment. The $K_{oc}$ values provided in Table 12 were obtained as described in Chapter 2.1.1.3. $K_{oc}$ values for carbamazepine (80 L/kg), sitagliptin (760 L/kg) and trimethoprim (900 L/kg) were extracted from the scientific literature (Löffler et al. 2005, Emea 2009, Straub 2013).

Table 12: Compartment specific half-lives of selected substances as calculated in Chapter 2.1.2.

| | $DegT_{50,ts}$ (days) | $DegT_{50,w}$ (days) | $DegT_{50,sed}$ (days) | $K_{oc}$ (L/kg) |
|---|---|---|---|---|
| Substance API9 | 18.2 | 1040 | 3 | 710 |
| Substance API8 | 201 | 19000 | 237 | 1930 |
| Substance API13 | 16.9 | 1200 | 0.6 | 30 |
| Substance API6 | 36.9 | 730 | 2.3 | 30 |

#### 4.2.1.2 Rhine monitoring study

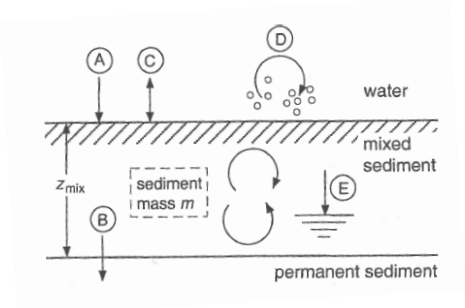
Ruff et al. (Ruff et al. 2015) followed the flow of the river Rhine and investigated the contamination by polar organic pollutants. A travelling water mass was sampled using weekly flow-proportional composite samples at ten different downstream sites. The first sampling site was located before the inlet of lake Constance (Diepoldsau). The sampling scheme also included main tributaries such as the rivers Aare, Neckar, Mosel and Main and the smaller tributaries Thur and Glatt. The water samples were analyzed using an analytical method based on solid phase extraction and high-resolution mass spectrometry. The analytical method targeted more than 300 substances. Samples of Lake Constance (upper Rhine) showed the presence of 83 substances of those target substances, whereas at the last sampling station, Bimmen, at the Dutch–German border, 143 substances were detected. The study was carried out for the period from 28.02 – 26.03.2011.

#### 4.2.1.3 Predictive modeling of concentrations in Rhine for selected pharmaceuticals

The applied model in this study is based on the spatial analysis of the Rhine river catchment underlying a dataset from (Ingold et al. in prep., Moser et al. in prep. ). The Rhine catchment area was divided into 18791 sub catchments. Stream gauge data of 932 gauging stations of the river network within the catchment area was used to calculate the discharge of the Rhine and its tributaries on an hourly basis for the year 2011. Further geographical location and person equivalent (PE) data of 2647 wastewater treatment plants (WWTPs) was collected within the catchment area. Correction factors were applied to the PE dataset, to include missing or small scale WWTPs within the dataset. In a next step, hydraulic parameters for the Rhine river such as water depth, Strickler coefficient, slope, river bed width, water level, river bed elevation and cross sectional perimeter were extracted from a GIS data base and implemented into AQUASIM, a software package for the identification and simulation of aquatic systems (Reichert 1994).

For the Rhine model, a river section compartment module was programmed within AQUASIM, with an explicit description of the water-sediment exchange of chemicals. Processes relevant for the water-sediment exchange were modeled according to the Surface Mixed Sediment Layer model (SMSL) as illustrated in Figure 19 (processes illustrated: A=particle settling, B=transfer into permanent sediment C=diffuse exchange, D=resuspension, E=chemical or biochemical degradation). Model parameters and processes were mostly chosen as described elsewhere (Fenner et al. 2002) and listed in Appendix 3 and Appendix 4 .

Figure 19: SMSL model, extracted from (Schwarzenbach et al. 2005)




For the river section compartment module, lateral and upstream water inflow and substance input were distinguished in AQUASIM. To differentiate lateral and upstream inputs, it is necessary to divide the stream of interest into river compartments. For the Rhine model, we defined 13 river compartments. The starting and end points of the compartments were chosen such that they were placed at the inflow locations of the 11 tributaries with the highest accumulated PE from WWTPs. Figure 20 depicts the first two river compartments in the Rhine model. Discharge and loads from smaller tributaries, e.g., the rivers Thur, Töss and Glatt in the case of the first compartment shown in Figure 20, are summed and distributed equally across the length of the entire river compartment. The water inflow and loads from the 11 largest tributaries (in terms of accumulated PE), such as the river Aare in Figure 20, are used as upstream inputs within the model, i.e., those inputs are routed to the starting point of the respective compartment. As illustrated in Figure 21, the largest tributaries therefore define the beginning of new river compartments within the Rhine model.

Figure 20: Illustration of river compartment implementation in AQUASIM

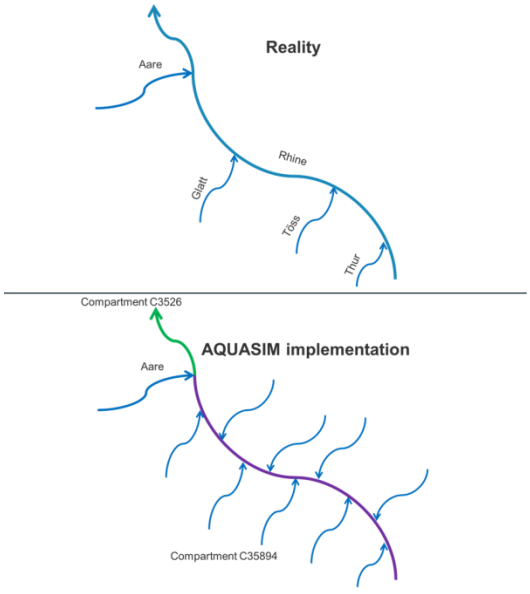
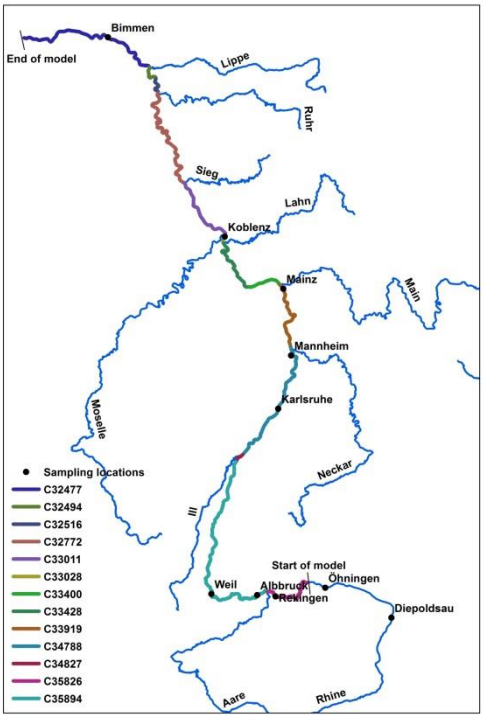


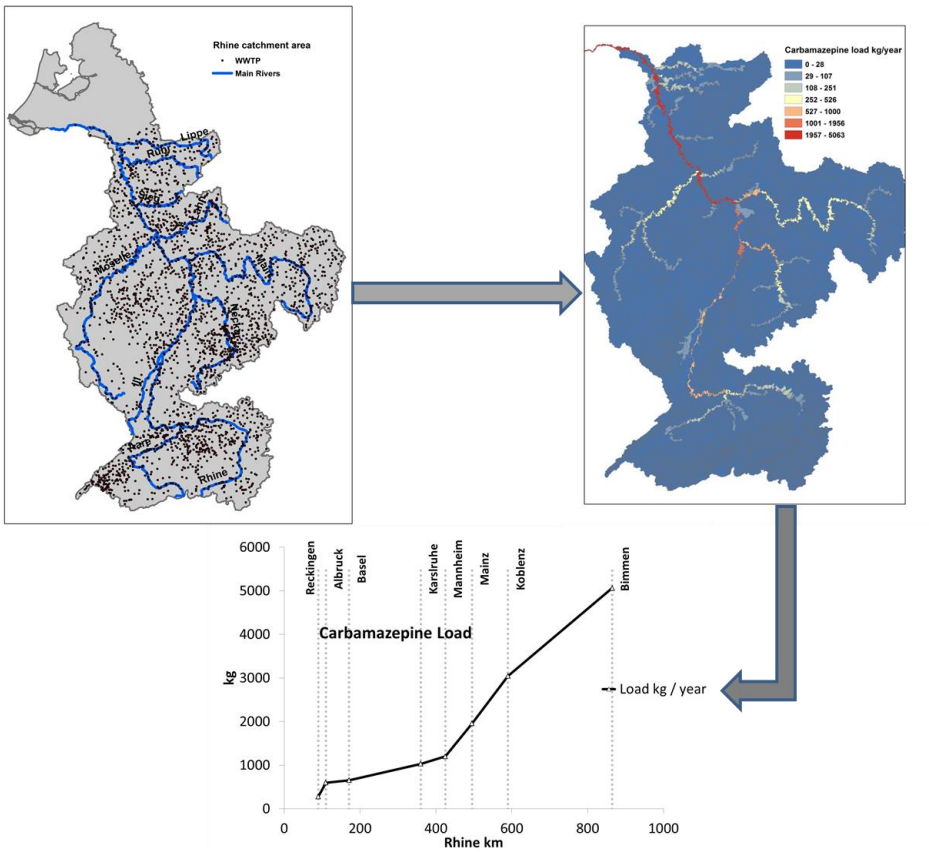

Figure 21 shows all river compartments as well as sampling locations described in (Ruff et al. 2015). The starting point of the model is located around 20 km behind the outlet of Lake Constance and the total length of the model is 974 km.

Figure 21: River compartments within the Rhine model




The substance input to the Rhine model was calculated for each sub catchment of the Rhine catchment area. Consumption, elimination and excretion rates of selected substances (obtained from (Singer et al. 2016)) were multiplied with the PE for each WWTP within the Rhine catchment area. Subsequently, the loads for each sub catchment were calculated. This was done by summing up in downstream direction the calculated substance loads from each WWTPs, resulting in a “load map” as exemplary illustrated in Figure 22 for carbamazepine.

Figure 22: Distribution of WWTPs in the Rhine catchment and illustration of the resulting load map for carbamazepine




In a next step, the lateral and upstream substance loads were extracted from the load map according to the pre-defined river compartments and entered as input into the Rhine model. Since consumption data, and excretion and removal factors were known to be highly uncertain, the input loads were adjusted using the measured concentrations for the tributaries as reported in (Ruff et al. 2015) to increase model accuracy. The adjustment was conducted in four steps:

1. For the rivers Aare, Neckar, Main and Mosel, the measured substance loads as reported in (Ruff et al. 2015) were used as input into the model instead of the calculated loads.

2. A correction factor was calibrated and applied to the other upstream inputs in the model. This correction factor was calculated from the fold-difference between measured loads reported in (Ruff et al. 2015) for the main Rhine tributaries (Aare, Neckar, Main and Mosel) and the loads calculated in the Rhine model as outline in the previous. Different correction factors were calculated for tributaries located in Switzerland (Aare) and tributaries located in Germany (Neckar, Main, Mosel) because consumption might vary between countries for legal and cultural reasons. The final correction factors employed are given in Table 13.

3. For the first river compartment, the upstream input was assumed to equal the substance load measured at the Öhningen sampling station. If for this station no concentration was measured, this value was estimated such as to fit the measured concentrations in Reckingen.

4. The lateral inputs of the first compartment are mainly constituted by the rivers Glatt, Thur and Töss. Within the Rhine model, these tributaries are categorized as lateral tributaries for the first compartment. The measured loads for the rivers Glatt and Thur were added up with the load for the river Töss

as extracted from the load map. The resulting load was applied as lateral input in AQUASIM for the first compartment.

The effectively calculated input loads are listed in Appendix 5 and Appendix 6. The simulations in AQUASIM were conducted for the time period of interest (28.02 – 26.03.2011).

A burn-in period of about 100 hours was used in the Rhine model to ensure that the model reached steady-state prior to the period of interest. Concentrations were predicted for Rhine sampling locations as investigated in (Ruff et al. 2015). The concentrations were calculated on an hourly basis, which allowed extracting a composite sample for the same time period as in (Ruff et al. 2015) for the sampled locations. The exact time frames for which samples were extracted are listed in Appendix 8.

#### 4.2.1.4 Availability of input parameters for chemical substances

Table 13 shows the availability of input parameters for the substances of interest in the Rhine model. In the first column, the availability of measured concentrations for the main tributaries Aare (A), Neckar (N), Main (Ma) and Mosel (M) is stated. The second column states whether a measured load for Öhningen was available, or whether the input load was approximated such as to fit the measured concentration at the first modeled sampling location in Reckingen. The lateral input of the first compartment is mainly constituted by the rivers Glatt (G), Thur (T) and Töss, of which G and T were sampled in (Ruff et al. 2015). For substance API13, the concentration in the river Thur was approximated as LOQ/2. The fifth column of Table 13 gives information on the availability of consumption data for Germany (G), Switzerland (CH) and France (F). If no data for Switzerland or France was available, German consumption data was used as a default. The correction factors that were applied to upstream inputs to the Rhine model are also listed in Table 13. Finally, in the last column, an overall confidence rating is provided for each substance based on data availability and the value of the correction factor, where 1 indicates high confidence and 3 low confidence in the model predictions.

Table 13: Additional information on substance-specific input parameters

| | Main tributaries (Ruff et al. 2015) | Input first compartment upstream (Ruff et al. 2015) | Input first compartment lateral (Ruff et al. 2015) | Consumption data (Singer et al. 2016) | Applied correction factor | Confidence |
|---|---|---|---|---|---|---|
| carbamazepine | A, N, Ma, Mo | Öhningen | Gl, Th | G, CH , F | 0.47 | 1 |
| Substance API9 | A, N, Ma, Mo | Öhningen | Gl, Th | G, CH , F | 0.7 | 1 |
| Substance API8 | A, N, Ma, Mo | Öhningen | Gl, Th | G, CH | 0.8 | 1 |
| Substance API13 | N, Mo | Approximated | Gl, Th (1/2 LOQ) | G, CH | 1.44 | 2 |
| Substance API6 | A, N, Ma, Mo | Approximated | Gl, Th | G, CH | 0.11 | 3 |
| sitagliptin | A, N, Ma, Mo | Approximated | Gl, Th | G | 0.84 | 3 |
| trimethoprim | A, N, Ma, Mo | Approximated | Gl, Th | G, CH , F | 0.1 | 3 |

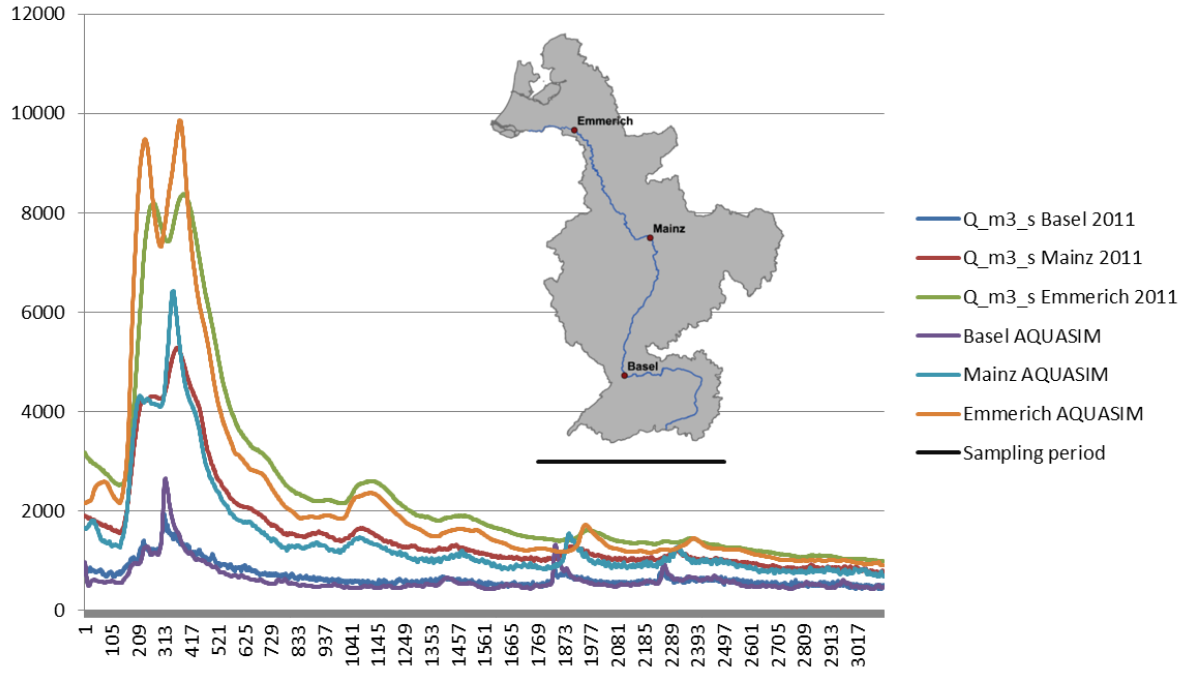
Abbreviations: A=Aare, N=Neckar, Ma=Main, Mo=Mosel; Gl=Glatt, Th=Thur; G=Germany, CH= Switzerland, F=France

## 4.2.2 Results

### 4.2.2.1 Accuracy of hydrological predictions

Figure 23 shows the discharge of the Rhine as computed in AQUASIM and actually measured for the year 2011 from January until April, highlighting the sampling period stated in (Ruff et al. 2015). The illustrated gauging stations are located in the upper Rhine (Basel), the middle Rhine (Mainz) and the lower Rhine (Emmerich), the corresponding discharge data was extracted from (Ingold et al. in prep., Moser et al. in prep. ). It is visible that in periods of high discharge the discharge computed by the program is slightly higher than the actually measured discharge, whereas for low discharge periods the predictions tend to underestimate the actual discharge. Overall, the computed discharge shows a strong correspondence with the measured discharge with an average relative deviation of – 2% relative to the measured data. Further, the discharge data of the Rhine for the different sampling locations as stated in (Ruff et al. 2015) was compared to the computed discharge data as extracted from AQUASIM for the time period of interest (Appendix 9). The average relative deviation of computed and measured discharge was -1 % relative to the measured data. The agreement between modeled and measured discharge was therefore considered sufficiently good for the model to be suitable for the prediction of substance concentrations.

Figure 23: Comparison of measured discharge with predicted discharge (AQUASIM)




### 4.2.2.2 External validation of consumption data and applied correction factors

To validate consumption data extracted from (Singer et al. 2016) and correction factors applied in this study as described in Chapter 4.2.1.3, a comparison with in- and outflow loads of WWTPs in Baden-Württemberg was conducted as given in (LUBW 2014). To validate the applied correction factors, the measured outflow loads of the substances API9, API6 and carbamazepine were compared with calculated WWTP outflow loads based on consumption, excretion and elimination data obtained from (Singer et al. 2016) for the selected substances.

The results of this validation as presented in Table 14 indicate that the procedure for the derivation of the correction factor works very well for the conservative substance carbamazepine, and also seems to

be still fairly consistent with measured WWTP outflow data for API9. For API6, in contrast, the analysis indicates that the low correction factor based on the measured tributary loads stands in clear conflict with the measured WWTP outflow loads that are even higher than predicted, and hence is very likely to reflect degradation in the tributaries rather than just input load correction.

Table 14: Comparison of correction factors used in this study with the ratio of the average measured WWTP outflow loads in Baden-Württemberg as given in (LUBW 2014) and the uncorrected outflow loads as predicted in this study according to the data given in (Singer et al. 2016).

| | Correction factors used in this study | (LUBW 2014) / (Singer et al. 2016) |
|---|---|---|
| **carbamazepin** | 0.47 | 0.53 |
| **API9** | 0.70 | 0.30 |
| **API6** | 0.11 | 4.68 |

#### 4.2.2.3 Comparison of predicted and measured substance concentrations along the river Rhine

In the following, for each group of substances as defined in Chapter 4.2.1.1, predicted and measured substance concentrations along the river Rhine are compared. In all figures, a 15% uncertainty interval is indicated for the measured concentrations as reported in (Ruff et al. 2015). If OECD 308 data was available for the selected substance, three degradation scenarios are presented using the degradation half-lives summarized in Table 12:

NO DEG: Degradation rate constants were set to 0 in the Rhine model.

DEG: Separate, compartment-specific rate constants as derived from the compartment-specific half-lives were used for the water and sediment compartment of the Rhine model.

DEG SYS : The total system degradation rate constant as derived from $DegT_{50,ts}$ was applied both to the water and sediment compartment of the Rhine model.

**Carbamazepine (substance group I)**

Figure 24 shows the predicted concentrations for carbamazepine and compares them to measured concentrations in the Rhine. The first scenario given (i.e., “loadmap”) used input loads extracted exclusively from the load map. One can see that using load map inputs only, modeled concentrations in the Rhine are on average by about a factor of 2.4 higher than measured concentrations. The conservative substance carbamazepine was therefore used to calibrate the correction factor methodology as described in Chapter 4.2.1.3. With the final input corrections as described in Chapter 4.2.1.3 applied, the average error for the predicted concentrations of carbamazepine was reduced to only 3 % relative to the measured data.

Figure 24: Comparison of predicted and measured concentrations for carbamazepine

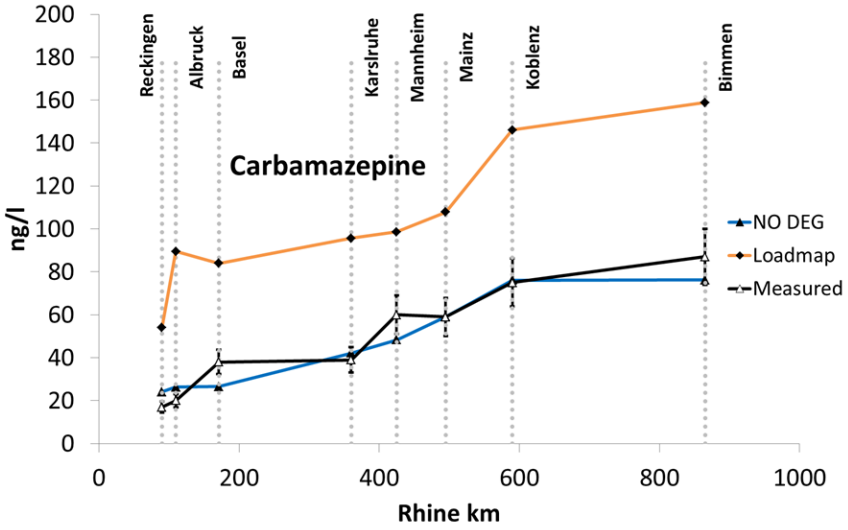
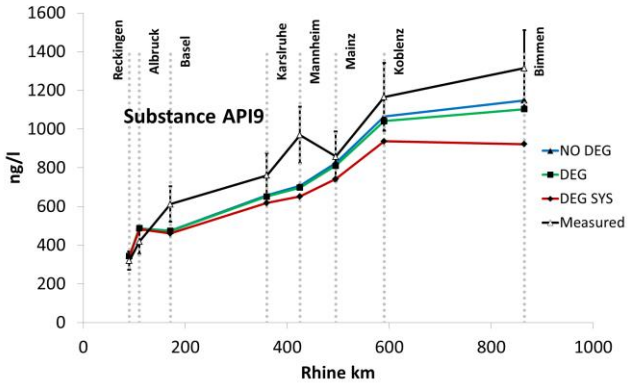



Figure 25 shows the predicted concentrations for the three scenarios “NO DEG”, “DEG” and “DEG SYS” for substances API8, API9, and API13 and compares them to measured concentrations in the Rhine. For substance API13, the measured concentration in Reckingen was approximated by dividing the level of quantification (LOQ) by 2 (indicated by orange data point).

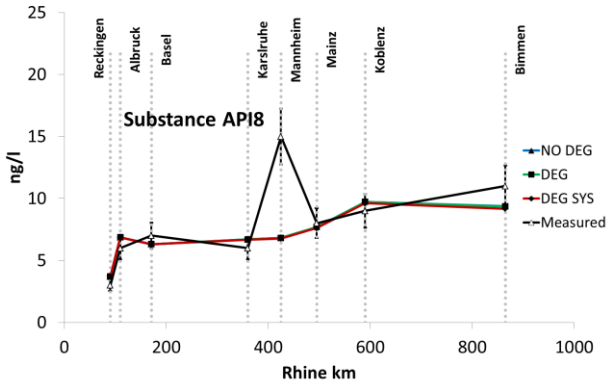
For all three substances, the scenarios “NO DEG” and “DEG” are mostly within the 15% measurement uncertainty of the Rhine monitoring study (Ruff et al. 2015) and can therefore be considered consistent with measured concentrations in the Rhine. The concentration of substance API8 in Mannheim was considered an outliers as it clearly deviated from both the measured and modeled trend along the Rhine. It can further be seen that for substances API9 and API13 the application of $DegT_{50,ts}$ in the “DEG SYS” scenario is overestimating degradation in the Rhine. This is not the case for substance API8, which has a $DegT_{50,ts}$ of 200.9 days (Table 12), which greatly exceeds the travel time of the investigated water parcel in the Rhine. Overall, one can see that predictions applying the compartments-specific degradation rate constants (“DEG”) are not significantly better or worse than scenarios with no degradation and only result in minor degradation of the selected substances in the Rhine for the period of interest.

Figure 25: Comparison of predicted and measured concentrations for substances API9, API8 and API13

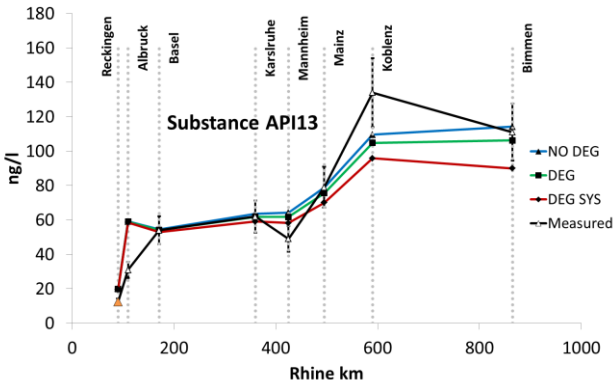
a)




b)




c)




**Conclusions**

Predictions using compartment-specific degradation half-lives ($DegT_{50,w}$, $DegT_{50,sed}$) are in accordance with measured concentrations in the Rhine.

Instead, the application of the total system half-life ($DegT_{50,ts}$) as degradation half-life in the Rhine overestimates degradation. This demonstrates that $DegT_{50,ts}$ as directly observed in the OECD 308 test system is not transferable to field situations with other water-sediment ratios.

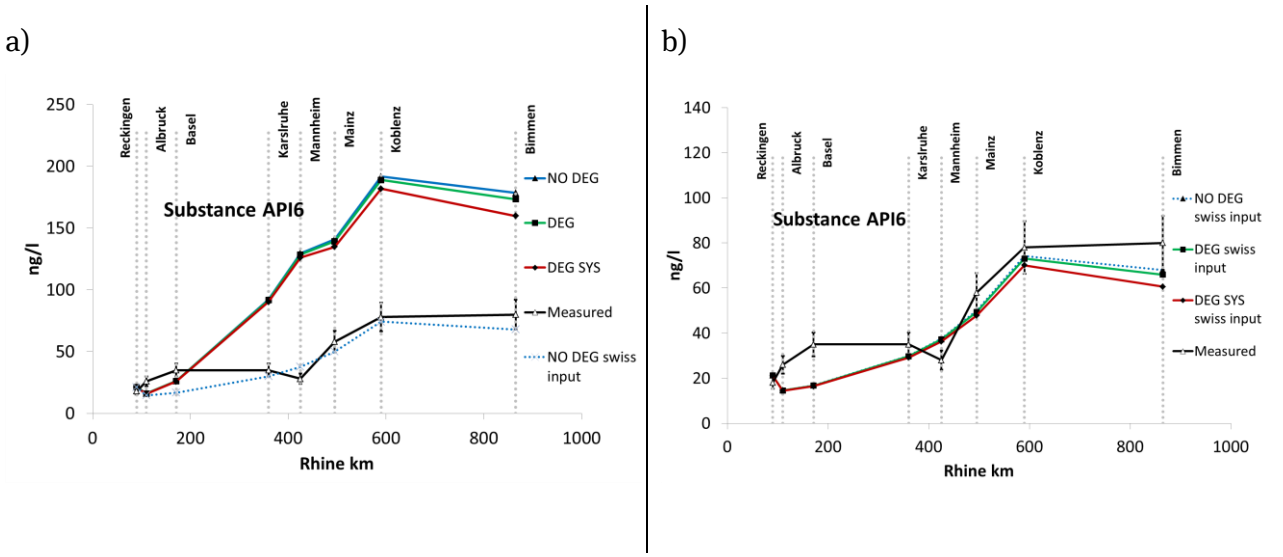
### Substance API6 (substance group II)

Figure 26a shows a significant discrepancy of modeled and measured concentrations for substance API6 in the Rhine for all three standard scenarios ("NO DEG", "DEG", "DEG SYS"). This deviation could potentially be due to three different reasons: (i) Either degradation is significantly stronger than suggested by the OECD 308 results, which would be supported by the analysis given in Chapter 4.2.2.2, or (ii) inappropriate consumption data were used. Interestingly, according to (Singer et al. 2016), the consumption in Germany for this substance is five times higher than in Switzerland. Since concentrations in Germany are clearly overestimated, a new scenario was run ("NO DEG swiss input") which

assumed Swiss consumption data for substance API6 also for Germany. Since this yielded a much better agreement with measured concentrations, degradation rate constants were applied to this consumption scenario as shown in Figure 26 b. However, none of the scenarios in Figure 26 b gives truly superior results. This might indicate that the problem is rather related to degradation being more extensive than suggested by the results of the OECD 308 study, which could point towards photolytic transformation. This is supported by the fact that literature-reported field studies carried out during different seasons for this substance point towards strongly decreased persistence during the spring months as compared to the winter months. Since degradation in the tributaries is currently not explicitly considered in the model, fast degradation in the tributaries can, however, not be included as possible mechanism with the current Rhine model.

Overall, the example of substance API6 illustrates the difficulty of evaluating degradation rate constants in the light of significant uncertainty about the spatial distribution of substance input. This, in turn, highlights the need for measured WWTP outflow data in the studied catchment if degradation rate constants are to be estimated based on concentration data measured in the field.

Figure 26: Comparison of predicted and measured concentrations for substance API6 with default input scenario (a) and “swiss input” scenario (b)



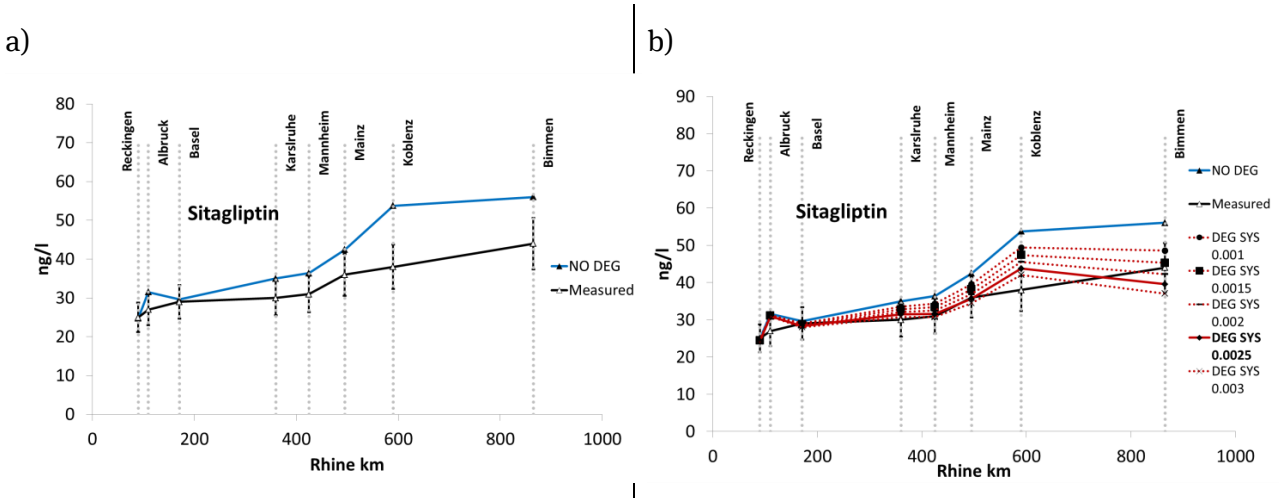

**Sitagliptin and trimethoprim (substance group III)**

For the identification of substances that are assumed to degrade well, the ratio of measured concentrations in Bimmen and Reckingen was analyzed. For the highly conservative substance carbamazepine, a concentration ratio of 5.1 (Bimmen/Reckingen) was calculated. Substances in Group B all showed ratios >3 (i.e., 4.1 (API9), 3.7 (API8), 8.9 (API13), 4.4 (API6)) assumed to be caused by accumulation in the Rhine and little degradation. The selected substances sitagliptin and trimethoprim showed measured concentration ratios of < 2 (Bimmen/Reckingen) and generally displayed a shallow slope of their concentration patterns along the Rhine. This led to the assumption that these substances undergo increased degradation in the Rhine river.

Figure 27a shows the modeled and measured concentration pattern for the antidiabetic drug sitagliptin along the Rhine. It can be seen that, with increasing travel distance along the Rhine, the discrepancy between measured and predicted concentrations, assuming no degradation, grows. This supports the hypothesis that sitagliptin is efficiently degraded in the Rhine. Therefore we used the

Rhine model to identify degradation rate constants that best predicted the measured concentrations (Figure 27 b). To this end, rate constants were varied between 0.001 and 0.003 $h^{-1}$ in increments of 0.0005 $h^{-1}$. The best fitting scenario according to residual errors was obtained with a rate constant of 0.0025$^{-1}$ applied to both the water and sediment compartment. This rate constant corresponds to a half-life of 11.6 days.

Figure 27: Comparison of predicted and measured concentrations for sitagliptin



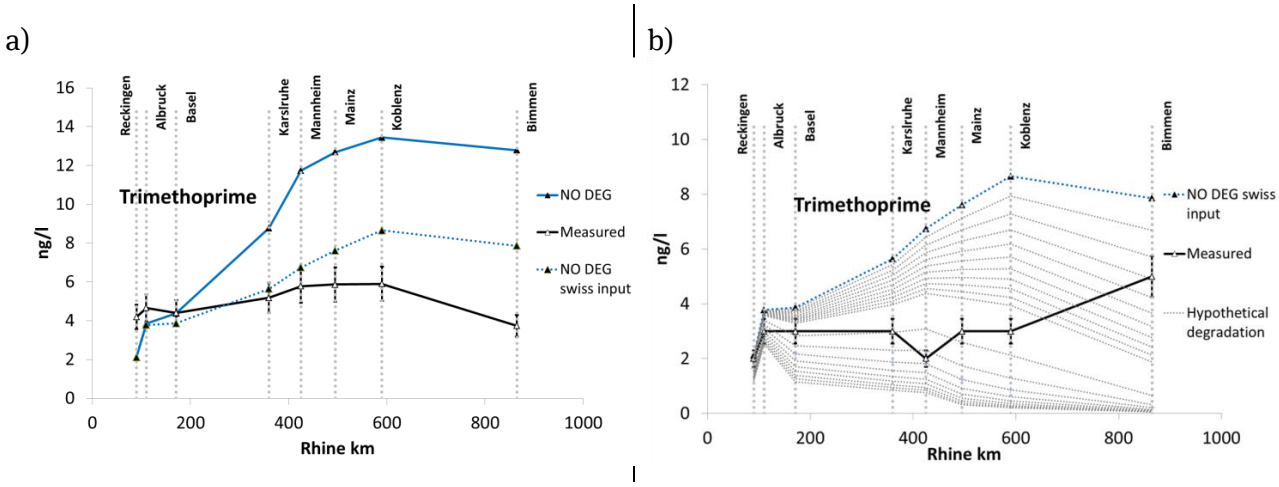

## Conclusion

For a substance with accurate consumption data, the comparison of predicted and measured concentrations can be used to estimate the substance's degradation half-life in the Rhine river.

The concentration pattern for the antibiotic drug trimethoprim (Figure 28a) obtained with the default Rhine model shows a similar discrepancy with measured concentrations as it was observed for substance API6. The yearly per capita consumption for trimethoprim in Germany was three times higher than in Switzerland according to (Singer et al. 2016). To decrease the deviation of modeled and measured concentrations, we used the same strategy as for API6 and applied the Swiss consumption estimate also to the German part of the river Rhine. This scenario is given as "NO DEG swiss input" in Figure 28a and serves as an initial scenario in Figure 28 b. The dashed lines in Figure 28 b illustrate how concentration predictions change when degradation rate constants from k=0.001 $h^{-1}$ to k= 0. 1 $h^{-1}$ are applied to the Swiss input scenario (in increments of 0.001 $h^{-1}$ from k=0.001 $h^{-1}$ to k=0.01 $h^{-1}$ and 0.01 $h^{-1}$ from k=0.01 $h^{-1}$ to k=0.1 $h^{-1}$ ) However, it seems that none of the degradation rate constants results in a concentration pattern that satisfactory reflects the measured concentration pattern, particularly with respect to the concentrations in Bimmen.

Since trimethoprim is also used as a veterinary pharmaceutical, the uncertainty with respect to the extent and spatial pattern of its input loads into the river Rhine is probably even higher than for the human pharmaceuticals discussed so far. Overall, the example of trimethoprim confirms the previous conclusion that uncertainty about the spatial distribution of substance input impedes confidently estimating a substance's degradation half-life in the Rhine river.

Figure 28: AQUASIM output for trimethoprim




**Conclusion**

Uncertainty about the spatial distribution of substance input impedes estimation of degradation half-lives in the river Rhine, and hence also impedes assessing the appropriateness of OECD 308 data to correctly reflect substance degradation.

#### 4.2.2.4 Evaluating the effect of degradation and sorption in concentration patterns in the river Rhine

In order to better understand under what circumstances degradation is sufficiently high to lead to observable deviation from the "NO DEG" case, the model was run with different degradation rate constants and with different assumptions as to where degradation takes place. To simulate substance input, the carbamazepine loads were used. The scenarios differentiated degradation in the total system (a), degradation only in the water phase of the system (b) and degradation only in the sediment phase of the river system (c). Degradation rate constants applied were 0.001, 0.005, 0.01, 0.05, 0.1 and 0.5 $h^{-1}$.

From comparing Figure 30a and b one can clearly see that application of degradation rate constants to the total system (i.e., water and sediment compartment) or to water only results in almost identical predicted concentrations. In contrast, if the same degradation rate constants are applied to the sediment only, the impact on the predicted concentrations is minor (Figure 30 c). A significant depletion of substance concentrations along the Rhine is only predicted when a sediment degradation rate constant of k=0.5 $h^{-1}$ is applied, which translates into a very low half-life of 0.06 days.

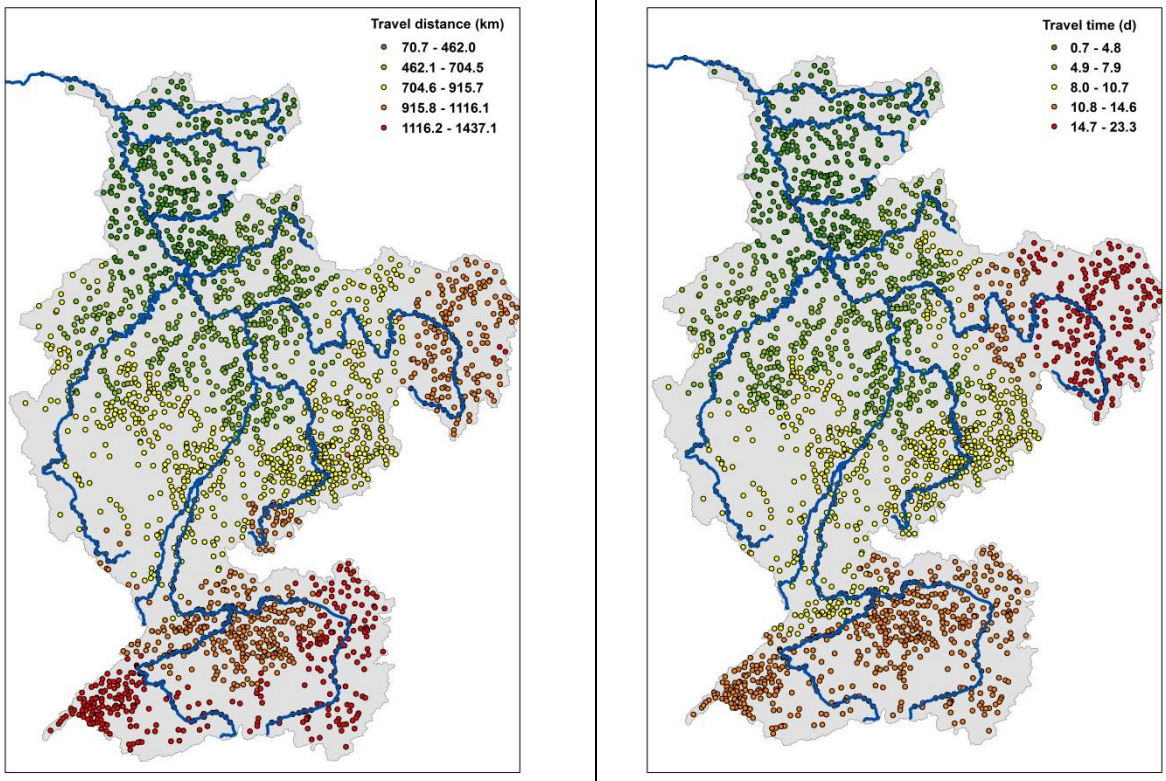
These simulations demonstrate two important points:

First, in a large stream like the river Rhine with an average depth of the water column of about 2.4 meters (Figure 31) and an assumed depth of the oxic sediment layer of 5 cm, degradation in the water column dominates the observed overall degradation in the river. For that reason, total system degradation half-lives obtained from an OECD 308 experiment, whose water-sediment ratio of 3-4:1 is about a factor of 50 lower than in the Rhine, cannot be directly transfer to a large stream like the Rhine.

Second, one can now delineate what kind of water half-lives would result in an observable degradation along the river Rhine. According to Figure 30a and b, degradation seems to become clearly distinct from measurement uncertainty somewhere in the range of degradation rate constants of >0.001-0.005

$h^{-1}$, corresponding to degradation half-lives in the range of <6-29 days. This is in agreement with the average travel time of all WWTP input loads along the river Rhine, as illustrated in Figure 29, which is approximately 7.7 days. The applied calculation methodology of the average travel time is described in Appendix 10.

Figure 29: Travel time and distance of WWTP input loads




Only compounds that degrade in the water column with a half-life in the range of the average travel time of all WWTP input loads can be expected to show a clearly recognizable degradation signature. As demonstrated in Chapter 2.1.2.2, such half-lives are rarely achieved due to biotransformation in the water column, and are only likely for substances that either show appreciable abiotic hydrolysis or photodegradation.

Figure 30: Hypothetical degradation scenarios differentiated for water, sediment and total system, assuming different degradation rates

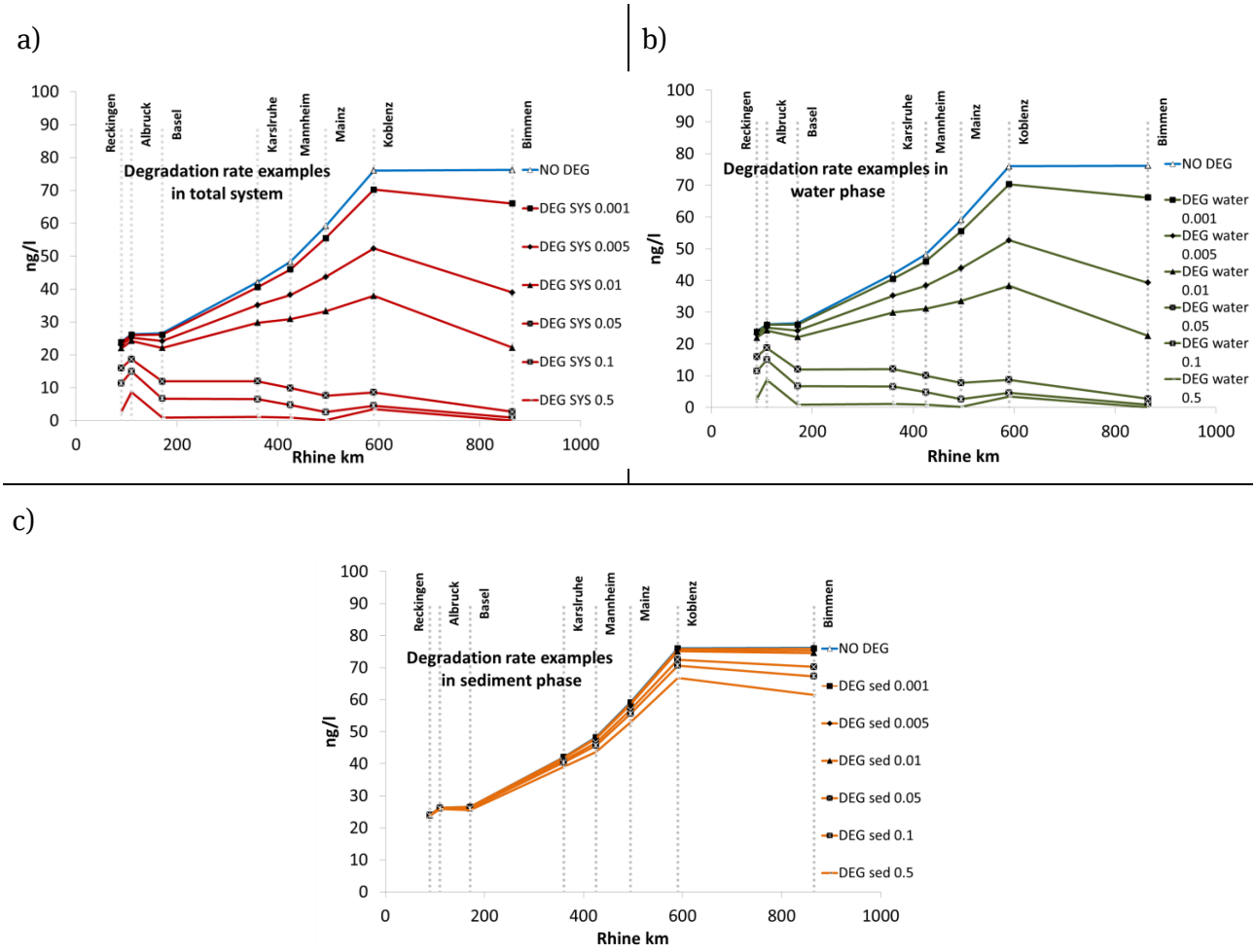
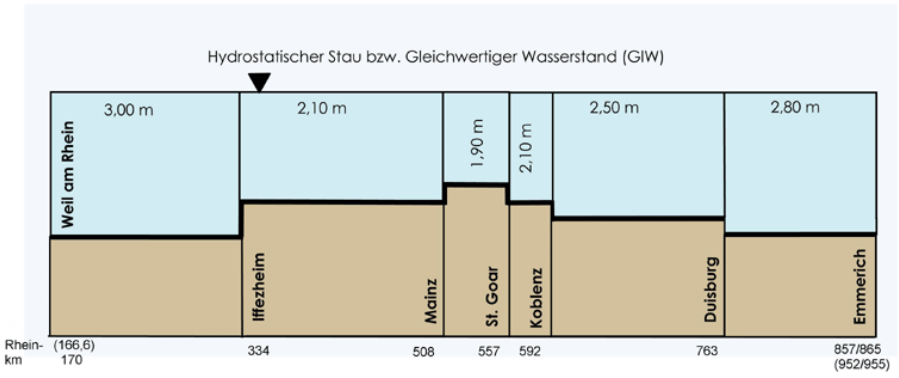


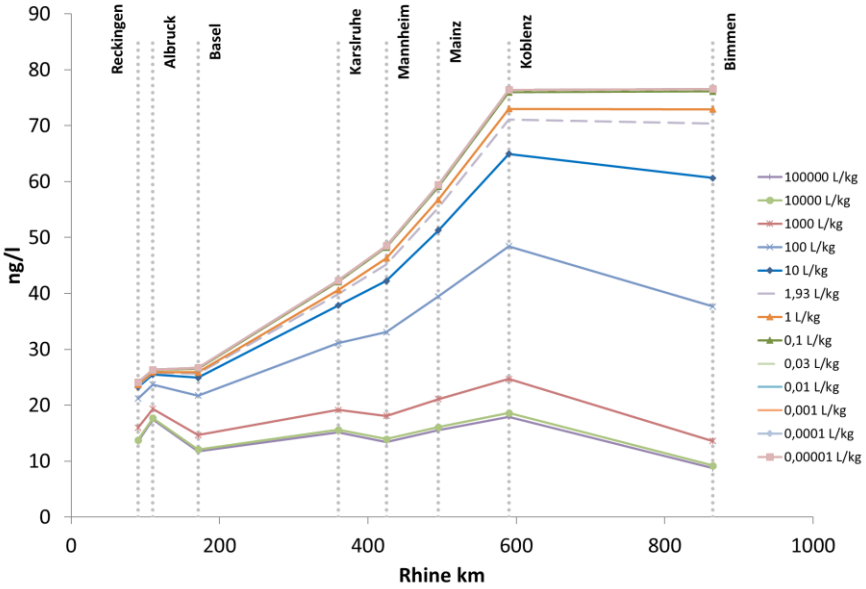

Figure 31: Average water level of the Rhine, extracted from (ELWIS 2014)




To investigate the influence of sorption on the concentration pattern along the Rhine, different hypothetical $K_{OC}$ values were applied to a hypothetical substance (carbamazepine input loads and no degradation assumed) as illustrated in Figure 32. As listed in Table 12, $K_{OC}$ values applied for substances used in this study vary from 30 L/kg (PH 13, PH 6) to 1930 L/kg (API8). Figure 32 shows that even the

highest $K_{OC}$ value of 1930 L/kg resulted in a difference of only 5% compared to the completely conservative case (0 L/kg). We can thus conclude that sorption to the sediment had very little influence on the fate of the substances modeled in this study.

Figure 32: Application of hypothetical $K_{OC}$ values using carbamazepine input loads. The value of 1930 L/kg indicates the highest $K_{OC}$ values of all seven compounds investigated in the Rhine model.




**Conclusions**

The water-sediment ratio as used in the OECD 308 guideline does not reflect situations in large rivers. Therefore, total system degradation half-lives obtained from OECD 308 experiments cannot be directly transferred to a large stream like the Rhine.

In large streams, only compounds that degrade in the water column with a half-life in the range of the average travel time of all WWTP input loads can be expected to show a clearly recognizable degradation signature. For the situation of the river Rhine, where WWTPs are distributed all along the length of the river, this results in a low average residence time of approximately 7.7 days. Such low degradation half-lives (< 10-20 d) are hardly achievable through biotransformation in the water column. As a consequence, only substances that either show appreciable abiotic hydrolysis or photodegradation are likely to be significantly degraded in the river Rhine.

## 4.3 Discussion and recommendations

One important finding from the Rhine study is that the considerable uncertainty about the level and spatial distribution of substance input into the Rhine makes the estimation of degradation half-lives from comparing measured and modeled concentrations rather difficult and prone to misinterpretation. Nevertheless, in combination with modeling of different scenarios regarding both input distribution and extent of degradation, several things could be learned from the Rhine study.

In large rivers such as the Rhine, for most except the most strongly sorbing substances, the bigger part of the substance mass resides in the water column. Therefore, the data obtained from OECD 308 studies, which are mostly representing biotransformation in the sediment, are not very relevant to representing degradation of chemicals in large rivers. Rather, results of abiotic hydrolysis studies, photodegradation studies, or OECD 309 would be considerably more informative to represent degradation of chemicals in large rivers. This said, if OECD 308 data are correctly interpreted as compartment-specific half-lives, which mostly results in high degradation half-lives in the water column, the results from OECD 308 experiments do not contradict the observations in the field. In the future, it would be interesting to obtain OECD 308 data for substances where the concentration patterns point towards a less conservative behavior in the river Rhine (e.g., sitagliptin, trimethoprim) to see whether the OECD 308 data do actually indicate fast degradation also in the water column. Similarly, monitoring data for strongly sorbing substances would complement the overall conclusions on the limited suitability of OECD 308 data to represent substance degradation in large streams.

For smaller rivers with significantly lower water levels on the order of < 0.5 m, more degradation can be expected to be observed in the overall river system, and OECD 308 might more appropriately reflect this situation. This is also suggested by the comparison of literature half-lives for laboratory and field systems as given in Table 11. The field sites in Table 11 all represent smaller rivers and it is found that laboratory and field total system half-lives never differ by more than a factor of 10.

This finally leads to the question what would be a good system to accurately observe degradation in the field. One option to reduce uncertainty in the input function, and hence increase the accuracy of estimating degradation, would be to study degradation in small streams with only one source of chemicals to the stream. (Schwientek et al. 2016), for instance, applied a Lagrangian sampling scheme to a 4 km river stretch with only minor discharge from tributaries and were able to identify the percent removal of a number of substances quite accurately. However, one drawback of studying degradation in small rivers is that the travel distances are typically rather small and that only rapid degradation is therefore observable at all. Alternatively, when studying larger river systems, outflow loads of representative WWTPs would need to be monitored in addition to concentrations in the river system to reduce uncertainty in the input function. An alternative observation system would be shallow lakes as demonstrated by the work of (Zou et al. 2015a). These have an increased residence time compared to flowing systems and often only one source of input also. However, it is unknown how well degradation observed in those systems represents degradation in flowing rivers. Finally, although unwanted, accidental industrial substance spills into large rivers, such as was the case during the accident in Schweizerhalle in 1986, would be another opportunity to determine degradation of the chemicals released in the spill. Today, temporally intermittent industrial inputs that are part of the regular production process such as, for instance, described in (Schlüsener et al. 2015), may present a similar opportunity.

# 5. Summary conclusions and recommendations

The aim of this research project was to assess the suitability of regulatory water-sediment simulation studies to identify persistent chemicals in surface waters, with a particular focus on rivers. Based on the results of this project, recommendations as to how to best make use of existing OECD 308 and OECD 309 data, and also how to potentially improve or complement interpretation of these simulation tests in the future should be given. To achieve this goal, the project had the following two objectives:

- Objective work package I: Evaluation and, if successful, establishment of an inverse modeling approach to derive different degradation half-lives and their uncertainty ranges for use as persistence indicators from OECD 308 and OECD 309 data for pesticides and pharmaceuticals.
- Objective work package II: Assessment of the representativeness of the laboratory-based OECD 308 and 309 simulation tests to reflect and predict chemicals' fate in actual surface water bodies.

In the following, the main conclusions from the project are summarized.

- *Derivation of total system half-life as persistence indicator:* The total system half-life ($DegT_{50,ts}$) is a meaningful yet system-specific degradation half-life due to the fact that compounds – except for the highly volatile ones – cannot escape the experimental systems unless transformed. It is the most robust among the degradation half-lives because it derives from directly observable and clearly identifiable data and requires fitting very simple models. It was found that the single-first order model (SFO) was the most robust kinetic model to calculate $DegT_{50,ts}$, giving sufficiently good fits to the observed data in all but one case. The more complicated DFOP, HS, and FOMC models were over-parameterised for most experiments, and therefore became numerically unstable for scarce data or outside the observed data range. Nevertheless, proper assessment of the model uncertainty indicated that $DegT_{50,ts}$ values had a significant uncertainty. For most cases, the relative uncertainty around the mean exceeded 20%, for about half of the cases it exceeded 100% (factor of two uncertainty).
- *Derivation of compartment-specific half-lives as persistence indicators:* $DissT_{50,w}$ and $DissT_{50,sed}$ mix up biotransformation and phase transfer and are therefore not recommended for persistence or exposure assessment. Instead, the compartment-specific half-lives $DegT_{50,w}$ and $DegT_{50,sed}$ are conceptually superior by indicating biotransformation separately, but require sophisticated inverse modeling for their derivation. Derivation of the bioavailability- and biomass-normalized second-order degradation rate constant $k'_{bio}$ is possible across different water-sediment systems and allows estimating $DegT_{50,w}$ and $DegT_{50,sed}$ based on $k'_{bio}$ in a rather universal way. However, $DegT_{50,w}$ and $DegT_{50,sed}$ values thus obtained remain very uncertain (factor of 10-100 uncertainty). The joint calibration of different experimental types at once via shared model parameters reduces the uncertainty of $k'_{bio}$, $DegT_{50,w}$ and $DegT_{50,sed}$, but this reduction of uncertainty is limited due to the limited information content of OECD 309 data.
- *Uncertainty considerations:* Uncertainty of persistence indicators is not acknowledged in current assessment practice. This leads to an incompatibility of the uncertain degradation half-lives ($DegT_{50,ts}$, $DegT_{50,w}$ and $DegT_{50,sed}$) and the presently used rigid persistence criteria.
- *Representativeness of OECD 308 and 309 simulation studies to reflect degradation in natural rivers (based on theoretical considerations):* Theoretical considerations of the different factors influencing biotransformation in water-sediment systems underline that there is no single test system that best represents the conditions in German river systems in general, nor might there even be a single test system that is most representative of a given type of river system in all aspects relevant to biotransformation. On similar grounds, it cannot be predicted whether any

of the two test systems would systematically yield more conservative results. Rather, the relative degree of conservatism of the two test systems would potentially be different for chemicals with strongly deviating properties (e.g., for strongly sorbing chemicals or for chemicals that are preferentially transformed under anaerobic conditions).

- *Representativeness of OECD 308 and 309 simulation studies to reflect degradation in natural rivers (based on river Rhine case study):* Predictions using compartment-specific degradation half-lives ($DegT_{50,w}$, $DegT_{50,sed}$) were found to not contradict measured concentrations in the Rhine. Instead, the application of the total system half-life ($DegT_{50,ts}$) as degradation half-life to both the water and sediment compartment of the Rhine model clearly overestimated degradation, demonstrating that $DegT_{50,ts}$ directly observed in the OECD 308 test system is not transferable to field situations with other water-sediment ratios. Specifically, in a large stream such as the Rhine, overall degradation of most substances (except for the most strongly sorbing ones) is dominated by degradation in the water. Therefore, the total system degradation half-life obtained from an OECD 308 experiment, whose water-sediment ratio of 3-4:1 is about a factor of 50 lower than in the Rhine, is not directly transferable. On the same grounds, for wastewater-borne substances with high to moderate polarity (i.e., $K_{oc}$ < 5000 L/kg), only those with half-lives in the water of <6-29 days, which is on the order of the average travel time of all WWTP input loads in the river Rhine catchment(i.e., approximately 7.7 days), can be expected to be significantly degraded in the Rhine. For most pesticides and pharmaceuticals studied, such half-lives were rarely achieved due to biotransformation in the water column. Rather substances either need to undergo appreciable abiotic hydrolysis or photodegradation to show significant degradation in the Rhine. As a consequence, data obtained from OECD 308 studies, which are mostly representing biotransformation in the sediment, are not very relevant to represent degradation of chemicals in large rivers.

Based on the above-summarized main conclusions, the following recommendations were derived in the project:

- *Data reporting requirements:* Deriving degradation half-lives of acceptable uncertainty and proper interpretation of the persistence indicators derived, requires a small set of crucial metadata to be reported (Table 7). The reporting of these metadata in a pre-defined format should be made an obligatory requirement for the study reports. Besides these crucial metadata, a precompiled residue table (Table 5) would help the assessment of data quality in terms of consolidated recovery rates and material balances.
- *Proper consideration of uncertainty:* Uncertainty assessment for all persistence indicators should become an integral part of persistence assessment and persistence classification procedures should be further developed to be able to account for this uncertainty in persistence assessment.
- *Proper consideration of environmental variability and variability in outcomes from different test systems:* Variability between test system outcomes is unavoidable and no clear association between test systems and environmental situations can be made. It is therefore strongly recommended that explicit and transparent strategies need to be implemented in the regulatory process to deal with these variabilities. Three different strategies to do so as outlined in the project are:
    - (i) Choosing the appropriate test system on a substance-by-substance basis such that, based on substance properties and emission scenarios, it best represents the exposure compartment where most of the substance mass will reside. This is in line with the current recommendations in the REACH Information Requirements for strongly sorbing substances to prefer testing according to OECD 308 over OECD 309. For these strongly

sorbing substances (about $K_{OC}$ > 5000 L/kg in this study), $DegT_{50,ts}$ is actually very similar to $DegT_{50,sed}$, and the total system half-life is indeed a good indicator of persistence in sediment. However, for substances with intermediate sorption behavior, it is more difficult to choose the test system and appropriate persistence indicator based on these principles because their mass distribution may shift significantly between the sediment and water column, depending on the type of water body (e.g., water column height and sediment properties);

- (ii) Derivation of the biomass- and bioavailability-normalized biotransformation rate constant $k'_{bio}$ as a means of extracting more fundamental information on biotransformation that is less dependent on the actual test system geometry and water:sediment ratio of a given test system, and hence shows a reduced variability compared to the degradation half-lives from different test systems. Since no persistence criterion for $k'_{bio}$ itself exists, its usage in persistence assessment would require for it to be converted back into compartment-specific half-lives, e.g., by using a set of standard sediment and water conditions;
- (iii) Using the outcome of any test system such as OECD 308, OECD 309 or a variant thereof for persistence assessment, while giving explicit consideration to the uncertainty inherent in doing so. This could be achieved by introducing a "safety factor", either in the persistence criterion itself or by multiplying the test outcome prior to comparison to the persistence criterion.

- *Revised testing requirements:* Overall, based on the results of the project, the execution of two simulation studies to assess biotransformation in water-sediment systems is recommended. These should be an OECD 308 study and a 309 study with as much suspended sediment as allowed. Doing so allows derivation of $k'_{bio}$ as a fundamental indicator of the biotransformation potential of a given substance that gets rid of many of the test system-related differences in degradation half-lives. This in turn allows deducing compartment-specific half-life indicators with reduced uncertainty, and allows considering the actual system dimensions in the field during exposure modelling as demonstrated in the Rhine case study. Adoption of a modified OECD 309 with higher (1:100) sediment-water ratios would further increase the robustness of $k'_{bio}$ and hence persistence indicators derived therefrom. However, it would need to be understood that direct estimation of $DegT_{50,ts}$ from such a system for comparison to water persistence criteria would likely not be appropriate because of the "unnaturally" high suspended sediment concentrations.
- *Strategies for P monitoring in the field:* The lack of field measurements of persistence is another factor impeding the assessment of the environmental representativeness of laboratory-based simulation studies. It is therefore useful to ask what would be a good system to accurately observe degradation half-lives in the field and whether such an endeavor could profit from any ongoing monitoring activities. The results of the Rhine case study clearly show that uncertainty in the inputs needs to be reduced. Therefore, either monitoring of representative WWTP outflows is included as integral part into such field studies or studies are carried out in small streams with only one source of chemicals to the stream. However, one drawback of studying degradation in small rivers is that the travel distances are typically rather small and that only rapid degradation is therefore observable at all. An alternative observation system would be shallow lakes. These have an increased residence time compared to flowing systems. However, it is unknown how well degradation observed in those systems represents degradation in flowing rivers. Finally, although unwanted, accidental industrial substance spills into large rivers, such as was the case during the accident in Schweizerhalle in 1986, would be another oppor-

tunity to determine degradation of the chemicals released in the spill. Today, temporally intermittent industrial inputs that are part of the regular production process may present a similar opportunity.

A number of points were raised by the findings of this project, which could, however, not or only partially be addressed. These present potential opportunities for follow-up research:

- The results from inverse modeling of OECD 308 and OECD 309 (pelagic) data indicated that half-lives in the water column of the 308 study and in the OECD 309 study were long, to the effect that three quarter of all pesticides and pharmaceuticals studied exceeded the persistence criterion in water. Joint modeling further indicated that OECD 309 data contributed little information as to the extent of biotransformation. These findings raise the question how useful and efficient it is to carry out OECD 309 (especially pelagic) studies, even more since hydrolysis testing is a standard data requirement for most chemical classes and seems to yield very similar half-lives as OECD 309 (pelagic). While the latter needs to be confirmed for a sufficiently large set of substances, it suggests that running OECD 309 with as much suspended sediment as allowed would lead to a larger information gain with respect to biotransformation in water-sediment systems.
- In the Rhine case study, OECD 308 data was only available for four of the substances monitored and all of them were fairly polar and recalcitrant in the water column. It would be interesting to extend the comparison of measured and predicted concentrations in the Rhine to substances that sorb more strongly and/or degrade more readily in the water column. In this way, the findings could be generalized to a broader range of substance behavior.
- Even large river systems such as the Rhine have maximal travel times of wastewater packages on the order of 20 days. Substances that pass the persistence criterion in water of < 40 d would thus still be transported to a large extent into the sea. This raises the question what the persistence criterion in water should actually protect and whether it is low enough to sufficiently protect aquatic resources.
- More recently, the use of benchmarking based on a set of reference compounds with well-known environmental degradation behavior has been suggested as a more appropriate way to assess persistence of chemicals. Using those chemicals as a reference system and assessing the behavior of any chemical against those reference chemicals (rather than some fixed persistence criteria) would allow circumventing the need for lab-to-field extrapolation and would also allow for a more explicit treatment of uncertainty. However, the usefulness of this concept still needs to be thoroughly explored. In particular, a set of reference chemicals with diverse sorption and biotransformation behavior would need to be defined and results for these chemicals in laboratory-based test systems and in field systems would need to be obtained. This would allow answering the question whether the relative behavior of substances is sufficiently conserved across these different systems to serve as a reference system for P and not P chemicals in water-sediment systems.

# 6. References

Aarewasser (2009). Nachhaltiger Hochwasserschutz Aare Thun–Bern Bern, Tiefbauamt Kanton Bern.

Akaike, H. (1974). "A new look at the statistical model identification." IEEE Transactions on Automatic Control 19(6): 716-723.

AND, F. P. and I. BRANCH (1987). "Guidelines for Determining Environmental Chemistry and Fate of Pesticides."

Araujo, L., M. E. Troconis, M. B. Espina and A. Prieto (2014). "Persistence of Ibuprofen, Ketoprofen, Diclofenac and Clofibric Acid in Natural Waters." Journal of Environment and Human. 1: 32-38.

Araujo, L., N. Villa, N. Camargo, M. Bustos, T. García and A. de Jesus Prieto (2011). "Persistence of gemfibrozil, naproxen and mefenamic acid in natural waters." Environmental Chemistry Letters 9(1): 13-18.

Bahlmann, A., M. G. Weller, U. Panne and R. J. Schneider (2009). "Monitoring carbamazepine in surface and wastewaters by an immunoassay based on a monoclonal antibody." Analytical and bioanalytical chemistry 395(6): 1809-1820.

Battersby, N. S. (1990). "A Review of Biodegradation Kinetics in the Aquatic Environment." Chemosphere 21(10-11): 1243-1284.

Battin, T. J., K. Besemer, M. M. Bengtsson, A. M. Romani and A. I. Packmann (2016). "The ecology and biogeochemistry of stream biofilms." Nature Reviews Microbiology 14(4): 251-263.

Boethling, R., K. Fenner, P. Howard, G. Klecka, T. Madsen, J. R. Snape and M. J. Whelan (2009). "Environmental Persistence of Organic Pollutants: Guidance for Development and Review of POP Risk Profiles." Integrated Environmental Assessment and Management 5(4): ISSN 1551-3777(print)|1551-3793(electronic).

Boonstra, H., E. P. Reichman and P. J. van den Brink (2011). "Effects of the veterinary pharmaceutical ivermectin in indoor aquatic microcosms." Archives of environmental contamination and toxicology 60(1): 77-89.

Boreen, A. L., W. A. Arnold and K. McNeill (2003). "Photodegradation of pharmaceuticals in the aquatic environment: A review." Aquatic Sciences 65(4): 320-341.

Bowmer, T. and A. Leopold (2004). 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. CEFIC. Rotterdam.

Bromilow, R. H., R. F. De Carvalho, A. A. Evans and P. H. Nicholls (2006). "Behavior of pesticides in sediment/water systems in outdoor mesocosms." Journal of Environmental Science and Health Part B 41(1): 1-16.

Burrows, H. D., M. Canle, J. A. Santaballa and S. Steenken (2002). "Reaction pathways and mechanisms of photodegradation of pesticides." Journal of Photochemistry and Photobiology B-Biology 67(2): 71-108.

Buser, H.-R., T. Poiger and M. D. Müller (1998). "Occurrence and fate of the pharmaceutical drug diclofenac in surface waters: rapid photodegradation in a lake." Environmental Science & Technology 32(22): 3449-3456.

Caracciolo, A. B., L. Patrolecco, M. Di Lenola, A. Battaglia and P. Grenni (2012). "Degradation of Emerging Pollutants in Aquatic Ecosystems." Cchemical Engineering 28.

Clara, M., B. Strenn and N. Kreuzinger (2004). "Carbamazepine as a possible anthropogenic marker in the aquatic environment: investigations on the behaviour of Carbamazepine in wastewater treatment and during groundwater infiltration." Water Research 38(4): 947-954.

Cuwvo (1999). Wvo-vergunningverlening. Commissie integraal Waterbeheer. https://www.helpdeskwater.nl/publish/pages/503/ciw41999-05handboekwvo-vergunningverlening.pdf.

Davis, J. W., S. Gonsior, G. Marty and J. Ariano (2005). "The transformation of hexabromocyclododecane in aerobic and anaerobic soils and aquatic sediments." Water Research 39(6): 1075-1084.

ECETOC (2010). Workshop Report No. 17: Significance of Bound Residues in Environmental Risk Assessment, 14-15 October 2009, Brussels. Brussels, European Centre for Ecotoxicology and Toxicology of Chemicals.

ECETOC (2014). Special Report No. 18: Information to be considered in a weight-of-evidence-based PBT/vPvB assessment of chemicals. Brussels, European Centre for Ecotoxicology and Toxicology of Chemicals.

ECHA (2008). Guidance on information requirements and chemical safety assessment, Part B: Hazard assessment, version 1.1. Guidance for the implementation of REACH. Helsinki, Finland, European Chemicals Agency.

ECHA (2016a). Guidance on Information Requirements and Chemical Safety Assessment, Chapter R.7b: Endpoint specific guidance. Helsinki, Finland, European Chemicals Agency.

ECHA (2016b). Guidance on Information Requirements and Chemical Safety Assessment, Chapter R.11: PBT/vPvB assessment. Helsinki, Finland, European Chemicals Agency.

ECHA (2016c). Guidance on information requirements and chemical safety assessment, Chapter R.16: Environmental Exposure Estimation. Guidance for the implementation of REACH. Helsinki, Finland, European Chemicals Agency.

ELWIS. (2014). "Fahrrinnensituation auf dem Rhein." Retrieved 06.09.2016, from https://www.elwis.de/Service/Telematikanwendungen/Inland-ECDIS/Allgemeines/ARGO/fahrrinnensituation/.

EMA (2011). 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). EMA/CHMP/SWP/44609/2010.

Emea (2009). Assessment Report for JANUVIA. London, European Medicies Agencies.

EPA (1998). "Fate, Transport and Transformation Test Guidelines OPPTS 835.3180 Sediment/Water Microcosm Biodegradation Test."

Ericson, J. F. (2007). "An evaluation of the OECD 308 water/sediment systems for investigating the biodegradation of pharmaceuticals." Environmental Science & Technology 41: 5803-5811.

Ericson, J. F., R. M. Smith, G. Roberts, B. Hannah, B. Hoeger and J. Ryan (2014). "Experiences with the OECD 308 transformation test: A human pharmaceutical perspective." Integrated environmental assessment and management 10(1): 114 - 124.

European Food Safety Authority (2008). "Opinion on a request from EFSA related to the default Q10 value used to describe the temperature effect on transformation rates of pesticides in soil - Scientific Opinion of the Panel on Plant Protection Products and their Residues (PPR Panel)." EFSA Journal 6(1): 622-n/a.

Fenner, K., C. Kooijman, M. Scheringer and K. Hungerbühler (2002). "Including transformation products into the risk assessment for chemicals: the case of nonylphenol ethoxylate usage in Switzerland." Environmental Science & Technology 36(6): 1147-1154.

Fenz, R., A. P. Blaschke, M. Clara, H. Kroiss, D. Mascher and M. Zessner (2005). "Quantification of sewer exfiltration using the anti-epileptic drug carbamazepine as marker species for wastewater." Water Science and Technology 52(9): 209-217.

Fischer, H., F. Kloep, S. Wilzcek and M. T. Pusch (2005). "A River's Liver – Microbial Processes within the Hyporheic Zone of a Large Lowland River." Biogeochemistry 76(2): 349-371.

FOCUS (1997). Surface water models and EU registration of plant protection products. 24.2.97. Report of the work of the Regulatory Modelling Working Group on Surface Water Models of FOCUS.

FOCUS (2006). Guidance Document in Estimating Persistence and Degradation Kinetics from Environmental Fate Studies in Pesticides in EU Registration. EC Document Reference Sanco/10058/2005 version 2.0,. Report of the FOCUS work group on degradation kinetics: 434.

FOCUS (2014). Generic guidance for Estimating Persistence and Degradation Kinetics from Environmental Fate Studies on Pesticides in EU Registration: 440.

Fono, L. J., E. P. Kolodziej and D. L. Sedlak (2006). "Attenuation of wastewater-derived contaminants in an effluent-dominated river." Environmental Science & Technology 40(23): 7257-7262.

Geweke, J. and H. Tanizaki (2003). "Note on the Sampling Distribution for the Metropolis-Hastings Algorithm." Communications in Statistics - Theory and Methods 32(4): 775-789.

Hand, L. H. and R. G. Oliver (2010). "The behavior of isopyrazam in aquatic ecosystems: Implementation of a tiered investigation." Environmental Toxicology and Chemistry 29(12): 2702-2712.

Honti, M. and K. Fenner (2015). "Deriving persistence indicators from regulatory water-sediment studies – Opportunities and limitations in OECD 308 data." Environmental Science & Technology 49 (10): 5879-5886.

Honti, M., S. Hahn, D. Hennecke, T. Junker, P. Shrestha and K. Fenner (2016). "Bridging across OECD 308 and 309 data in search of a robust biotransformation indicator." Environmental Science & Technology 50(13): 6865-6872.

Huntscha, S., H. Singer, S. Canonica, R. P. Schwarzenbach and K. Fenner (2008). "Input dynamics and fate in surface water of the herbicide metolachlor and of its highly mobile transformation product metolachlor ESA." Environmental Science & Technology 42(15): 5507-5513.

Ingold, K., H.-P. Bader, L. Herzog, F. Metz, A. Moser, R. Scheidegger and C. Stamm. "Misfit between physical affectedness and regulatory embeddedness of drinking water suppliers along the Rhine River." Global Environmental Change. (in prep.)

Jasper, J. T., Z. L. Jones, J. O. Sharp and D. L. Sedlak (2014). "Biotransformation of Trace Organic Contaminants in Open-Water Unit Process Treatment Wetlands." Environmental Science & Technology 48(9): 5136-5144.

Johansson, H., K. Jonsson, K. J. Forsman and A. Wörman (2001). "Retention of conservative and sorptive solutes in streams—simultaneous tracer experiments." Science of the total environment 266(1): 229-238.

Krause, W. J., W. Rost and T. Lüllwitz (2009). "Zur Bestimmung von Fließzeiten, Fließgeschwindigkeiten und longitudinaler Dispersion in der Mosel mit 3HHO als Leitstoff." Hydrologie und Wasserbewirtschaftung: 360-379.

Kunkel, U. and M. Radke (2008). "Biodegradation of acidic pharmaceuticals in bed sediments: Insight from a laboratory experiment." Environmental Science & Technology 42(19): 7273-7279.

Kunkel, U. and M. Radke (2011). "Reactive Tracer Test To Evaluate the Fate of Pharmaceuticals in Rivers." Environmental Science & Technology 45(15): 6296-6302.

Lam, M. W. and S. A. Mabury (2005). "Photodegradation of the pharmaceuticals atorvastatin, carbamazepine, levofloxacin, and sulfamethoxazole in natural waters." Aquatic Sciences 67(2): 177-188.

Lam, M. W., C. J. Young, R. A. Brain, D. J. Johnson, M. A. Hanson, C. J. Wilson, S. M. Richards, K. R. Solomon and S. A. Mabury (2004). "Aquatic persistence of eight pharmaceuticals in a microcosm study." Environmental Toxicology and Chemistry 23(6): 1431-1440.

LeFevre, G. H., C. E. Mueller, R. J. Lo, R. G. Luthy and E. S. Sattely (2015). "Rapid Phytotransformation of Benzotriazole Generates Synthetic Tryptophan and Auxin Analogs in Arabidopsis." Environmental Science & Technology 49(18): 10959-10968.

Li, Z., M. P. Maier and M. Radke (2014). "Screening for pharmaceutical transformation products formed in river sediment by combining ultrahigh performance liquid chromatography/high resolution mass spectrometry with a rapid data-processing method." Analytica Chimica Acta 810: 61-70.

Li, Z., A. Sobek and M. Radke (2015). "Flume Experiments To Investigate the Environmental Fate of Pharmaceuticals and Their Transformation Products in Streams." Environmental Science & Technology: 6009-6017.

Liber, K., K. R. Solomon and J. H. Carey (1997). "Persistence and fate of 2, 3, 4, 6-tetrachlorophenol and pentachlorophenol in limnocorrals." Environmental Toxicology and Chemistry 16(2): 293-305.

Lin, A. Y. C. and M. Reinhard (2005). "Photodegradation of common environmental pharmaceuticals and estrogens in river water." Environmental Toxicology and Chemistry 24(6): 1303-1309.

Lock, M. A., R. R. Wallace, J. W. Costerton, R. M. Ventullo and S. E. Charlton (1984). "River Epilithon: Toward a Structural-Functional Model." Oikos 42(1): 10-22.

Löffler, D., J. Römbke, M. Meller and T. A. Ternes (2005). "Environmental fate of pharmaceuticals in water/sediment systems." Environmental Science & Technology 39(14): 5209-5218.

LUBW (2014). Spurenstoffinventar der Fließgewässer in Baden-Württemberg. Karlsruhe, Ministerium für Umwelt, Klima und Energiewirtschaft Baden-Württemberg Stuttgart.

Lyon, D. R. and S. E. Ziegler (2009). "Carbon cycling within epilithic biofilm communities across a nutrient gradient of headwater streams." Limnology and Oceanography 54(2): 439-449.

Matamoros, V., A. Duhec, J. Albaigés and J. M. Bayona (2008). "Photodegradation of Carbamazepine, Ibuprofen, Ketoprofen and 17α-Ethinylestradiol in Fresh and Seawater." Water, Air, and Soil Pollution 196(1): 161-168.

Matthies, M., K. Solomon, M. Vighi, A. Gilman and J. V. Tarazona (2016). "The origin and evolution of assessment criteria for persistent, bioaccumulative and toxic (PBT) chemicals and persistent organic pollutants (POPs)." Environmental Science: Processes & Impacts 18(9): 1114-1128.

Mechteld, H. and A. Koelmans (2016). "Analyzing the Limitations and the Applicability Domain of Water–Sediment Transformation Tests like OECD 308." Environmental Science & Technology 50(19): 10335-10342.

Moser, A., H.-P. Bader, F. Fenicia, M. Honti, R. Scheidegger, D. Wemyss and C. Stamm. "Spatio-temporal patterns of micropollutants in the Rhine basin. ." Hydrology and Earth System Sciences. (in prep. )

Nakada, N., K. Kiri, H. Shinohara, A. Harada, K. Kuroda, S. Takizawa and H. Takada (2008). "Evaluation of pharmaceuticals and personal care products as water-soluble molecular markers of sewage." Environmental Science & Technology 42(17): 6347-6353.

Nelder, J. A. and R. Mead (1965). "A Simplex method for function minimization." The computer Journal 7(4): 308-313.

Niu, J., L. Zhang, Y. Li, J. Zhao, S. Lv and K. Xiao (2013). "Effects of environmental factors on sulfamethoxazole photodegradation under simulated sunlight irradiation: Kinetics and mechanism." Journal of Environmental Sciences 25(6): 1098-1106.

OECD (2002). Test No. 308: Aerobic and Anaerobic Transformation in Aquatic Sediment Systems.

OECD (2004). Test No. 309: Aerobic Mineralisation in Surface Water – Simulation Biodegradation Test.

OECD (2006). Test No. 53: Guidance document on simulated freshwater lentic field tests (outdoor microcosms and mesocosms).

Packer, J. L., J. J. Werner, D. E. Latch, K. McNeill and W. A. Arnold (2003). "Photochemical fate of pharmaceuticals in the environment: Naproxen, diclofenac, clofibric acid, and ibuprofen." Aquatic Sciences 65(4): 342-351.

Pottgiesser and Sommerhäuser (2008). "Beschreibung und Bewertung der deutschen Fließgewässertypen - Steckbriefe und Anhang." https://www.umweltbundesamt.de/sites/default/files/medien/378/publikationen/texte_43_2014_hydromorphologische_steckbriefe_der_deutschen_fliessgewaesssertypen_0.pdf

Radke, M., C. Lauwigi, G. Heinkele, T. E. Mürdter and M. Letzel (2009). "Fate of the antibiotic sulfamethoxazole and its two major human metabolites in a water sediment test." Environmental Science & Technology 43(9): 3135-3141.

Radke, M. and M. P. Maier (2011). "Screening des Verhaltens von Arzneimittelwirkstoffen in einem Wasser-Sediment-Testsystem." Mitteilungen der Fachgruppe Umweltchemie und Ökotoxikologie 17(3): 54-56.

Radke, M. and M. P. Maier (2014). "Lessons learned from water/sediment-testing of pharmaceuticals." Water Research 55: 63-73.

Radke, M., H. Ulrich, C. Wurm and U. Kunkel (2010). "Dynamics and Attenuation of Acidic Pharmaceuticals along a River Stretch." Environmental Science & Technology 44(8): 2968-2974.

Reichert, P. (1994). "AQUASIM–a tool for simulation and data analysis of aquatic systems." Water Science and Technology 30(2): 21-30.

Rua-Gomez, P. C. and W. Puettmann (2013). "Degradation of lidocaine, tramadol, venlafaxine and the metabolites O-desmethyltramadol and O-desmethylvenlafaxine in surface waters." Chemosphere 90(6): 1952-1959.

Ruff, M., M. S. Mueller, M. Loos and H. P. Singer (2015). "Quantitative target and systematic non-target analysis of polar organic micro-pollutants along the river Rhine using high-resolution mass-spectrometry–Identification of unknown sources and compounds." Water Research 87: 145-154.

Sanderson, H., B. Laird, L. Pope, R. Brain, C. Wilson, D. Johnson, G. Bryning, A. S. Peregrine, A. Boxall and K. Solomon (2007). "Assessment of the environmental fate and effects of ivermectin in aquatic mesocosms." Aquatic Toxicology 85(4): 229-240.

Schlüsener, M. P., U. Kunkel and T. A. Ternes (2015). "Quaternary triphenylphosphonium compounds: A new class of environmental pollutants." Environmental science & technology 49(24): 14282-14291.

Schrader, K. K., C. S. Tucker, M. Q. Regt and S. K. Kingsbury (2000). "Evaluation of limnocorrals for studying the effects of phytotoxic compounds on plankton and water chemistry in aquaculture ponds." Journal of the World Aquaculture Society 31(3): 403-415.

Schwarzenbach, R. P., B. I. Escher, K. Fenner, T. B. Hofstetter, C. A. Johnson, U. von Gunten and B. Wehrli (2006). "The challenge of micropollutants in aquatic systems." Science 313(5790): 1072-1077.

Schwarzenbach, R. P., P. M. Gschwend and D. M. Imboden (2005). Environmental Organic Chemistry, John Wiley & Sons.

Schwientek, M., G. Guillet, H. Ruegner, B. Kuch and P. Grathwohl (2016). "A high-precision sampling scheme to assess persistence and transport characteristics of micropollutants in rivers." Science of the Total Environment 540: 444-454.

SETAC (1995). Procedures for Assessing the Environmental Fate & Ecotoxicity of Pesticides. ISBN: 978-1-880611-76-0. Society of Environmental Toxicology and Chemistry.

Shrestha, P., T. Junker, K. Fenner, S. Hahn, M. Honti, R. Bakkour, C. Diaz and D. Hennecke (2016). "Simulation Studies to Explore Biodegradation in Water–Sediment Systems: From OECD 308 to OECD 309." Environmental Science & Technology 50(13): 6856-6864.

Singer, H. P., A. E. Wössner, C. S. McArdell and K. Fenner (2016). "Rapid Screening for Exposure to “Non-Target” Pharmaceuticals from Wastewater Effluents by Combining HRMS-Based Suspect Screening and Exposure Modeling." Environmental Science & Technology 50 (13): 6698-6707

Solomon, K., J. Yoo, D. Lean, N. Kaushik, K. Day and G. Stephenson (1985). "Dissipation of permethrin in limnocorrals." Canadian Journal of Fisheries and Aquatic Sciences 42(1): 70-76.

Straub, J. O. (2013). "An environmental risk assessment for human-use trimethoprim in European surface waters." Antibiotics 2(1): 115-162.

Thomas, K. A. and L. H. Hand (2011). "Assessing the potential for algae and macrophytes to degrade crop protection products in aquatic ecosystems." Environmental Toxicology and Chemistry 30(3): 622-631.

Thomas, K. A. and L. H. Hand (2012). "Assessing the metabolic potential of phototrophic communities in surface water environments: Fludioxonil as a model compound." Environmental Toxicology and Chemistry 31(9): 2138-2146.

Tixier, C., H. P. Singer, S. Canonica and S. R. Muller (2002). "Phototransformation of triclosan in surface waters: A relevant elimination process for this widely used biocide - Laboratory studies, field measurements, and modeling." Environmental Science & Technology 36(16): 3482-3489.

Tixier, C., H. P. Singer, S. Oellers and S. R. Muller (2003). "Occurrence and fate of carbamazepine, clofibric acid, diclofenac, ibuprofen, ketoprofen, and naproxen in surface waters." Environmental Science & Technology 37(6): 1061-1068.

UBA (2016). Confidential data, German Environmental Agency.

USEPA (1982). Pesticide Assessment Guidelines Subdivision N. Office of Pesticides and Toxic Substances. Washington.

VICH (2004). VICH-GL38. Environmental impact assessment (EIAS) for veterinary medicinal products - Phase II Guidance. Brussels, VICH International Cooperation on Harmonisation of Technical Requirements for Registration of Veterinary Medicinal Products: 36.

Vione, D., R. Das, F. Rubertelli, V. Maurino, C. Minero, S. Barbati and S. Chiron (2010). "Modelling the occurrence and reactivity of hydroxyl radicals in surface waters: implications for the fate of selected pesticides." International Journal of Environmental Analytical Chemistry 90(3-6): 260-275.

Wanner, O., T. Egli, T. Fleischmann, K. Lanz, P. Reichert and R. P. Schwarzenbach (1989). "Behavior of the insecticides disulfoton and thiometon in the Rhine River: a chemodynamic study." Environmental Science & Technology 23(10): 1232-1242.

Wells, G. F., H.-D. Park, B. Eggleston, C. A. Francis and C. S. Criddle (2011). "Fine-scale bacterial community dynamics and the taxa–time relationship within a full-scale activated sludge bioreactor." Water Research 45(17): 5476-5488.

WSV. (2016). "Pegel Online." http://pegelonline.wsv.de/gast/start, 01.06.2016.

Zou, H., M. MacLeod and M. S. McLachlan (2014). "Evaluation of the potential of benchmarking to facilitate the measurement of chemical persistence in lakes." Chemosphere 95: 301-309.

Zou, H., M. Radke, A. Kierkegaard, M. MacLeod and M. S. McLachlan (2015a). "Using Chemical Benchmarking to Determine the Persistence of Chemicals in a Swedish Lake." Environmental Science & Technology 49(3): 1646-1653.

Zou, H., M. Radke, A. Kierkegaard and M. S. McLachlan (2015b). "Temporal Variation of Chemical Persistence in a Swedish Lake Assessed by Benchmarking." Environmental Science & Technology 49(16): 9881-9888.

# 7. Appendix

Appendix 1: Degradation half-lives for pharmaceuticals reported in literature in water-sediment systems

| Literature Source | (Kunkel et al. 2008) | (Li et al. 2015) | (Löffler et al. 2005) | (Radke et al. 2009) | (Radke et al. 2014) | (Radke 2011) | (Kunkel et al. 2011) | (Zou et al. 2015a) | (Lam et al. 2004) | (Caracciolo et al. 2012) | (Araujo et al. 2011) | (Araujo et al. 2014) | (Radke et al. 2010) | (Ericson 2007) | (Boonstra et al. 2011) | (Sanderson et al. 2007) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| System type | Flume experiment | Flume experiment | OECD 308 | OECD 308 | Artificial River water | Artificial River water | Tracer | Chemical Benchmarking | Microcosm | Microcosm | Surface water only | Surface lake water | Tracer | OECD 308 | Microcosm | Mesocosm |
| bezafibrate | 2.5 - 4.3 | | | | | | | 14 | | | | | | | | |
| diclofenac | 3.2 - 8.5 | | | | 11.3 | | | 10 - 13 | | | | 2.14 ± 0.2 – 60 ± 9.4 | | | | |
| gemfibrozil | 5.6 | | | | | | | 47 | | | 119.5 ± 15.6 - 288.8 ± 61.3 | | | | | |
| ibuprofen | 1.2 - 2.5 | 1.8 | < 6 | | | 2.4 - 45 | 10 ± 1.3 h | | | | | 17.8 ±2.1 - 247.7 ± 30.1 | | | | |
| naproxen | 5.4 - 6.9 | | | | | | | | | <22 | 10.2 ± 0.5 - 14.6 ± 1.3 | | 3.6 ± 2.1 | | | |
| acetaminophen | | 1.8 | | | | | | | | | | | | | | |
| bezafibrate | | 1.9 | | | | | | | | | | | | | | |

| Literature Source | (Kunkel et al. 2008) | (Li et al. 2015) | (Löffler et al. 2005) | (Radke et al. 2009) | (Radke et al. 2014) | (Radke 2011) | (Kunkel et al. 2011) | (Zou et al. 2015a) | (Lam et al. 2004) | (Caracciolo et al. 2012) | (Araujo et al. 2011) | (Araujo et al. 2014) | (Radke et al. 2010) | (Ericson 2007) | (Boonstra et al. 2011) | (Sanderson et al. 2007) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| bicalutamide | | 37 - 49 | | | | | | | | | | | | | | |
| carbamaze-pine | | | | | | | | | 82 ± 11 | | | | | | | |
| chlorthalidone | | | | | | | | | | | | | | | | |
| clofibric acid | | 12-14 | | | | | 2.5 ± 0.5 | | | | | | | | | |
| fluconazole | | | | | | | | | | | | | | | | |
| furosemide | | 15 - 16 | | | | 15 - 52 | | <10 | | | | | | | | |
| hydrochloro-thiazide | | 57 | | | 52.4 | | | | | | | | | | | |
| ketoprofen | | 2 -2.2 | | | | | | <2 | | | | | | | | |
| metoprolol | | 5.5 - 7.2 | | | | 6.1 - 25 | | | | | | | | | | |
| naproxen | | 5.6 - 7.1 | | | | 4.8 - 17 | | | | | | | | | | |
| propranolol | | 5.6 - 7.1 | | | 11.1 | 3 - 14 | | | | | | | | | | |

| Literature Source | (Kunkel et al. 2008) | (Li et al. 2015) | (Löffler et al. 2005) | (Radke et al. 2009) | (Radke et al. 2014) | (Radke 2011) | (Kunkel et al. 2011) | (Zou et al. 2015a) | (Lam et al. 2004) | (Caracciolo et al. 2012) | (Araujo et al. 2011) | (Araujo et al. 2014) | (Radke et al. 2010) | (Ericson 2007) | (Boonstra et al. 2011) | (Sanderson et al. 2007) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| sotalol | | 6 - 9.4 | | | | | | | | | | | | | | |
| sulfamethoxazole | | 34 | | | | | | 26 | 19 ± 1.2 | | | | | | | |
| tramadol | | 49 | | | | | | | | | | | | | | |
| paracetamol | | | 3.1 ± 0.2 | | | | | | | | | | | | | |
| clofibric acid | | | 119 ± 7 | | | | | | | | | | | | | |
| diazepam | | | 311 ± 25 | | | | | | | | | | | | | |
| oxazepam | | | 54 ± 3 | | | | | | | | | | | | | |
| carbamazepine | | | 328 | | - | | | 1200 - 1400 | | | | | | | | |
| CBZ-diol | | | 8 | | | | | | | | | | | | | |
| iopromide | | | 29 ± 4 | | | | | | | | | | | | | |
| Sulfamethoxazole | | | | 8.5 - 17.2 | | | | | | | | | | | | |
| climbazole | | | | | | | | 18 | | | | | | | | |

| Literature Source | (Kunkel et al. 2008) | (Li et al. 2015) | (Löffler et al. 2005) | (Radke et al. 2009) | (Radke et al. 2014) | (Radke 2011) | (Kunkel et al. 2011) | (Zou et al. 2015a) | (Lam et al. 2004) | (Caracciolo et al. 2012) | (Araujo et al. 2011) | (Araujo et al. 2014) | (Radke et al. 2010) | (Ericson 2007) | (Boonstra et al. 2011) | (Sanderson et al. 2007) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| diatrizoic acid | | | | | | | | 131 - 216 | | | | | | | | |
| hydrochloro-thiazide | | | | | | | | 3 | | | | | | | | |
| acetamino-phen | | | | | | | | | 0.9 ± 0.2 | | | | | | | |
| atorvastatin | | | | | | | | | 6.6 ± 0.2 | | | | | | | |
| caffeine | | | | | | | | | 1.5 ± 0.4 | | | | | | | |
| levofloxacin | | | | | | | | | 5 ± 0.1 | | | | | | | |
| sertraline | | | | | | | | | 6.3 ± 0.2 | | | | | | | |
| trimethoprim | | | | | | | | | 5.7 ± 0.1 | | | | | | | |
| mefenamic acid | | | | | | | | | | | 15.5 ± 2.9 - 27.0 ± 6.6 | | | | | |
| exemestane | | | | | | | | | | | | | | 9.9 | | |
| varenicline | | | | | | | | | | | | | | 22 - 29 | | |

| Literature Source | (Kunkel et al. 2008) | (Li et al. 2015) | (Löffler et al. 2005) | (Radke et al. 2009) | (Radke et al. 2014) | (Radke 2011) | (Kunkel et al. 2011) | (Zou et al. 2015a) | (Lam et al. 2004) | (Caracciolo et al. 2012) | (Araujo et al. 2011) | (Araujo et al. 2014) | (Radke et al. 2010) | (Ericson 2007) | (Boonstra et al. 2011) | (Sanderson et al. 2007) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ivermectin | | | | | | | | | | | | | | | 1.1 - 8.3 | 3-5 |

Appendix 2: Photochemical half-lives of selected chemicals

| | Photo-chemical Half-life | Test system | Reference |
|---|---|---|---|
| diclofenac | 39 min | River Water | (Packer et al. 2003) |
| gemfibrozil | 15 h | River Water | (Lin et al. 2005) |
| ibuprofen | 205 h | River Water | (Lin et al. 2005) |
| naproxen | 42 min | River Water | (Packer et al. 2003) |
| clofibric acid | 50 h | River Water | (Packer et al. 2003) |
| ketoprofen | 2.4 min | River Water | (Matamoros et al. Matamoros2008 2008) |
| sulfamethoxazole | 39 / 50 min | Natural water, results for initial concentrations of 1mg/l and 5 mg/l | (Niu et al. 2013) |
| carbamazepine | 115 ± 4 h | Natural water | (Lam et al. 2005) |

Appendix 3: Model parameters as described in (Schwarzenbach et al. 2005)

| Parameter | Description | Unit | Expression |
|---|---|---|---|
| d_sed | Thickness of the sediment layer | m | 0.05 |
| fsed | % organic carbon in Sediment | | 0.005 |
| poc | Concentration of particulate organic matter in water | kg/m^3 | 0.002 |
| por | Porosity of sediment | | 0.8 |
| psed | Sediment density | kg/m^3 | 2500 |
| u_dep | Deposition flow | kg/m^2*h | 5.1*10^(-3)/24 |
| u_diff | Diffusion velocity | m/h | 0.16/24 |
| u_sed | Sedimentation velocity | m/h | 2/24 |
| resusp | Resuspension Flow | kg/m^2*h | 1*10^(-4)/24 |

Appendix 4: Model processes between water and sediment as described in (Fenner et al. 2002)

| Process description | Calculation |
|---|---|
| Fraction of substance adsorbing to particles in the water column | $frac_p^x = \frac{POC * K_{oc}^x}{1 + POC * K_{oc}^x}$ |
| Fraction of particles adsorbing to particles in sediment | $frac_{sed}^x = \frac{(1-por) * p^{sed} * f^{sed} * K_{oc}^x}{por + (1-por) * p^{sed} * f^{sed} * K_{oc}^x}$ |
| Total transport of substance from water into sediment | $u_{wsed}^x = \frac{u^{sed}}{h_w} * frac_p^x + \frac{u^{diff}}{h_w} * (1 - frac_p^x)$ |
| Total transport of substance from sediment into water | $u_{sedw}^x = \frac{U^{resusp}}{(1-por) * p^{sed} * h_{sed}} * frac_{sed}^x + \frac{u^{diff}}{h_{sed}} * (1 - frac_{sed}^x)$ |
| Burial to permanent sediment | $K_{out,sed} = \frac{U^{dep}}{(1-por) * p^{sed} * h_{sed}}$ |

Appendix 5: Lateral input loads (bold numbers are load map values that were modified according to described methodology in chapter 4.2.1.3)

| Lateral substance input in ng/h | carbamaze-pine | Substance API9 | Substance API8 | Substance API13 | Substance API6 | Substance API6 (swiss in-put only) | sitagliptin | trime-thoprim | trime-thoprim (swiss input only) |
|---|---|---|---|---|---|---|---|---|---|
| C32477 | 5.79E+11 | 6.39E+12 | 4.45E+10 | 3.16E+11 | 2.84E+12 | 6.21E+11 | 2.51E+11 | 2.45E+11 | 1.01E+11 |
| C32494 | 1.01E+10 | 1.09E+11 | 7.34E+08 | 5.44E+09 | 4.96E+10 | 1.04E+10 | 4.21E+09 | 4.46E+09 | 1.56E+09 |
| C32516 | 1.62E+09 | 1.75E+10 | 1.18E+08 | 8.74E+08 | 7.97E+09 | 1.68E+09 | 6.77E+08 | 7.17E+08 | 2.52E+08 |
| C32772 | 3.01E+10 | 3.25E+11 | 2.19E+09 | 1.62E+10 | 1.48E+11 | 3.12E+10 | 1.26E+10 | 1.33E+10 | 4.68E+09 |
| C33011 | 6.64E+09 | 7.18E+10 | 4.85E+08 | 3.59E+09 | 3.27E+10 | 6.89E+09 | 2.78E+09 | 2.95E+09 | 1.03E+09 |
| C33028 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| C33400 | 5.60E+09 | 6.06E+10 | 4.09E+08 | 3.03E+09 | 2.76E+10 | 5.81E+09 | 2.34E+09 | 2.49E+09 | 8.71E+08 |
| C33428 | 1.29E+08 | 1.39E+09 | 9.40E+06 | 6.96E+07 | 6.35E+08 | 1.34E+08 | 5.39E+07 | 5.71E+07 | 2.00E+07 |
| C33919 | 1.40E+10 | 1.51E+11 | 1.02E+09 | 7.54E+09 | 6.88E+10 | 1.45E+10 | 5.84E+09 | 6.19E+09 | 2.17E+09 |
| C34788 | 2.97E+10 | 3.24E+11 | 2.19E+09 | 1.62E+10 | 1.48E+11 | 3.12E+10 | 1.26E+10 | 1.26E+10 | 5.15E+09 |
| C34827 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| C35826 | **6.90E+09** | **9.33E+10** | **1.09E+09** | **3.02E+09** | **1.00E+10** | **1.00E+10** | **6.11E+09** | **1.01E+09** | **1.01E+09** |
| C35894 | 1.78E+10 | 2.01E+11 | 1.45E+09 | 9.84E+09 | 8.67E+10 | 1.99E+10 | 8.01E+09 | 7.16E+09 | 3.47E+09 |

Appendix 6: Upstream input loads (bold numbers were Loadmap values were modified according to described methodology in chapter 4.2.1.3)

| Upstream substance input in ng/h | carbamaze-pine | substance API9 | substance API8 | substance API13 | Substance API6 | Substance API6 (swiss in-put only) | sitagliptin | trime-thoprim | trime-thoprim (swiss input only) |
|---|---|---|---|---|---|---|---|---|---|
| C32477 | 1.05E+10 | 1.69E+11 | 1.30E+09 | 1.74E+10 | 1.21E+10 | 2.54E+09 | 7.84E+09 | 9.90E+09 | 3.47E+08 |
| C32494 | 1.54E+10 | 2.48E+11 | 1.91E+09 | 2.55E+10 | 1.78E+10 | 3.74E+09 | 1.15E+10 | 1.45E+10 | 5.10E+08 |
| C32516 | 1.07E+10 | 1.72E+11 | 1.32E+09 | 1.77E+10 | 1.23E+10 | 2.59E+09 | 7.98E+09 | 1.01E+10 | 3.53E+08 |
| C32772 | 5.14E+09 | 8.28E+10 | 6.39E+08 | 8.51E+09 | 5.93E+09 | 1.25E+09 | 3.85E+09 | 4.85E+09 | 1.70E+08 |
| C33011 | **2.22E+10** | **7.21E+11** | **2.68E+09** | **5.23E+10** | **1.98E+10** | **1.98E+10** | **2.36E+10** | **1.55E+09** | **1.55E+09** |
| C33028 | 1.42E+10 | 2.28E+11 | 1.76E+09 | 2.34E+10 | 1.63E+10 | 3.44E+09 | 1.06E+10 | 1.34E+10 | 4.68E+08 |
| C33400 | **8.03E+10** | **1.13E+12** | **1.09E+10** | **1.33E+11** | **1.21E+11** | **1.21E+11** | **5.63E+10** | **8.15E+09** | **8.15E+09** |
| C33428 | 7.64E+09 | 1.23E+11 | 9.49E+08 | 1.27E+10 | 8.82E+09 | 1.85E+09 | 5.71E+09 | 7.21E+09 | 2.53E+08 |
| C33919 | **5.04E+10** | **6.61E+11** | **6.07E+09** | **7.53E+10** | **4.98E+10** | **4.98E+10** | **3.55E+10** | **4.88E+09** | **4.88E+09** |
| C34788 | 2.33E+10 | 4.02E+11 | 3.19E+09 | 4.25E+10 | 2.96E+10 | 6.24E+09 | 1.92E+10 | 1.47E+10 | 1.47E+09 |
| C34827 | 1.28E+09 | 2.06E+10 | 1.59E+08 | 2.12E+09 | 1.48E+09 | 3.10E+08 | 9.57E+08 | 1.21E+09 | 4.23E+07 |
| C35826 | **8.33E+09** | **1.32E+11** | **1.31E+09** | **1.17E+10** | **1.73E+10** | **1.73E+10** | **1.07E+10** | **2.02E+09** | **2.02E+09** |
| C35894 | **2.18E+10** | **5.08E+11** | **8.27E+09** | **8.28E+10** | **3.04E+09** | **3.04E+09** | **3.11E+10** | **4.23E+09** | **4.23E+09** |

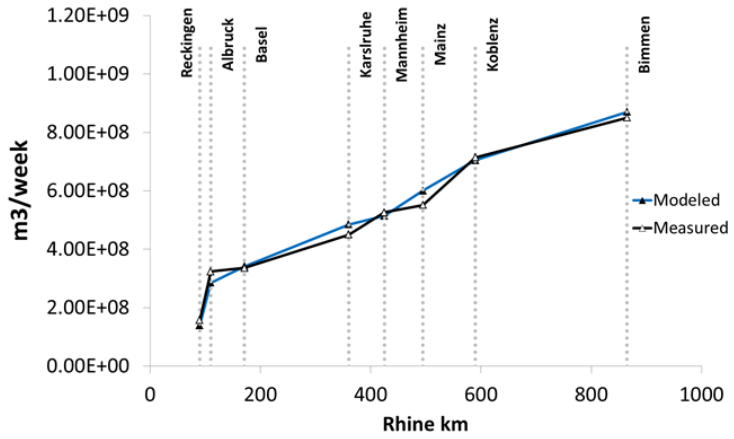
Appendix 7: Applied correction factors

| Substance | Correction Factor |
|---|---|
| carbamazepin | 0.47 |
| Substance API9 | 0.7 |
| Substance API8 | 0.8 |
| Substance API13 | 1.44 |
| Substance API6 | 0.11 |
| sitagliptin | 0.84 |
| trimethoprim | 0.1 |

Appendix 8: Time periods of AQUASIM sample extraction

| | Start day | End day | Start hour | End hour |
|---|---|---|---|---|
| Reckingen | 67 | 74 | 1608 | 1776 |
| Albruck | 67 | 74 | 1608 | 1776 |
| Basel | 69 | 77 | 1656 | 1848 |
| Karlsruhe | 73 | 80 | 1752 | 1920 |
| Mannheim | 74 | 81 | 1776 | 1946 |
| Mainz | 75 | 82 | 1800 | 1968 |
| Koblenz | 76 | 83 | 1824 | 1992 |
| Bimmen | 79 | 86 | 1896 | 2064 |

Appendix 9: Weekly discharge at sampling locations, modeled and extracted from (Ruff et al. 2015)




Appendix 10: Calculation methodology of average chemical travel time in the Rhine river

Based on the stream network in the Rhine catchment area, the Rhine catchment area was categorized into nine river basins with different flow velocities. The AQUASIM Rhine model was used to calculate average flow velocities of the Rhine river for the different river basins (alpine Rhine, high Rhine, upper Rhine, middle Rhine, low Rhine). As exemplarily illustrated for the high Rhine, the simulation was conducted for four time points throughout the year and then averaged (Figure 34). Average flow velocities of tributaries were either extracted and approximated from literature (Aarewasser 2009, Krause et al. 2009) or based on gauge data (WSV 2016).

Then, for each WWTP, the course its wastewater package traveled was delineated and the distance it traveled in any of the nine river basins was recorded. These distances were then converted into travel times using the flow velocities as described above, and summed up for each WWTP. Finally, those WWTP-specific travel times were averaged to an average travel time using the person equivalents (PE) of each WWTP as weighing factors.

Figure 33: River basins within the Rhine catchment area

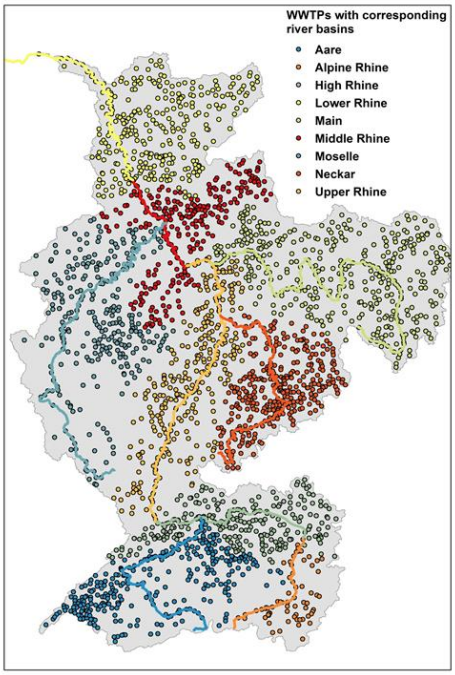
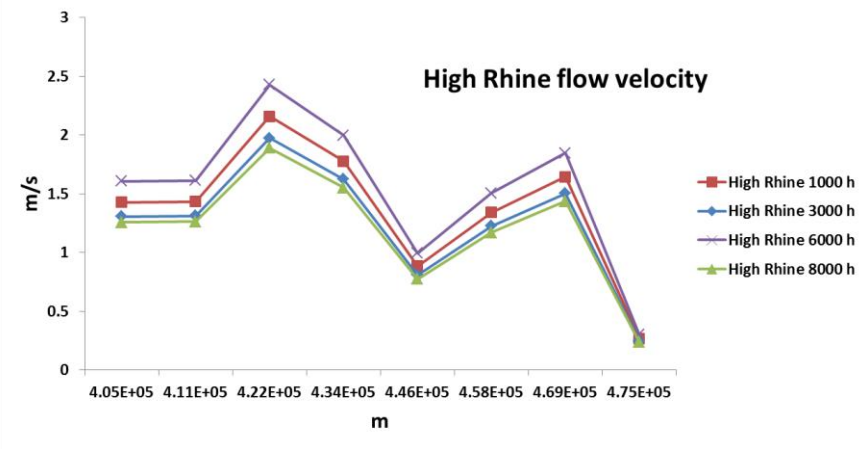



Figure 34: Exemplary AQUASIM output of flow velocity calculation for the high Rhine




Appendix 11: Electronic Supplement

Tables about compound properties, experimental metadata, and half-life quantiles from individual experiments and joint calibration.