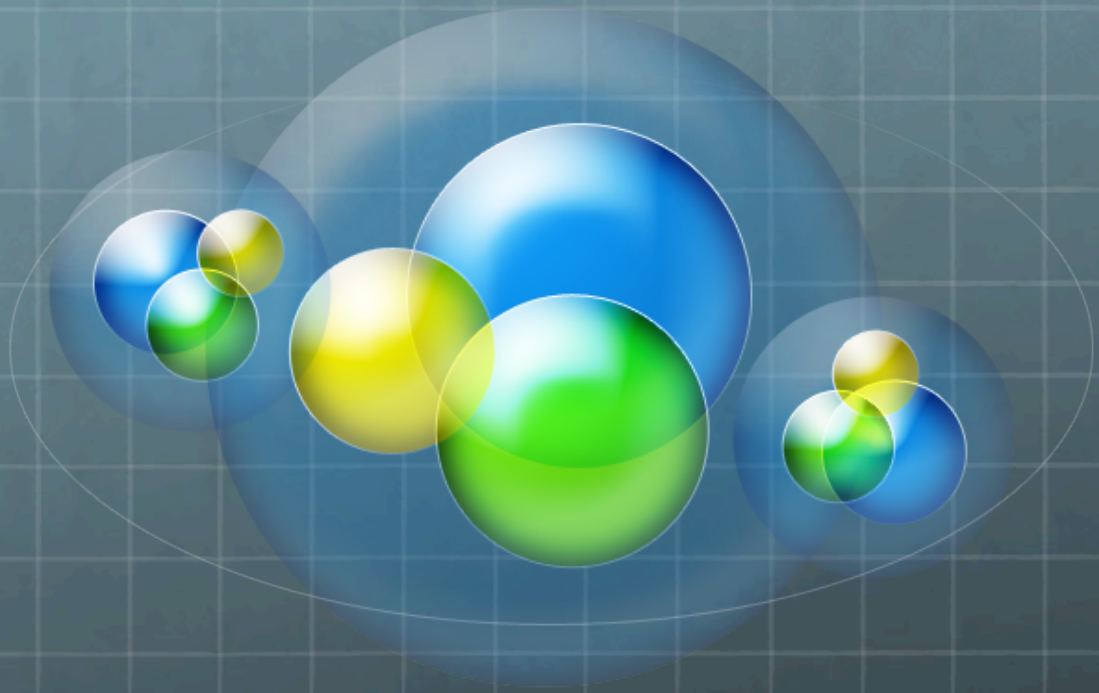
HelmholtzZentrum münchen
Deutsches Forschungszentrum für Gesundheit und Umwelt

CADASTER

MARIE CURIE ACTIONS



Environmental Risk Assessment

using the QSPR-Thesaurus Web Tool

P. Sopasakis, I. V. Tetko, P. Kunwar, S. Brandmaier, S. Novotarskyi, L. Charochkina, V. Prokopenko and W. J. G. M. Peijnenburg

Case studies on the development and application of *in silico* techniques for Environmental Hazard and Risk Assessment

OpenTox

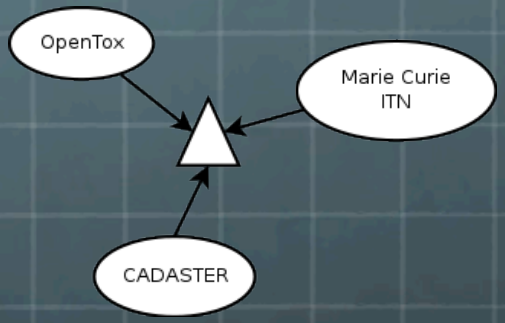
An open source predictive toxicology framework

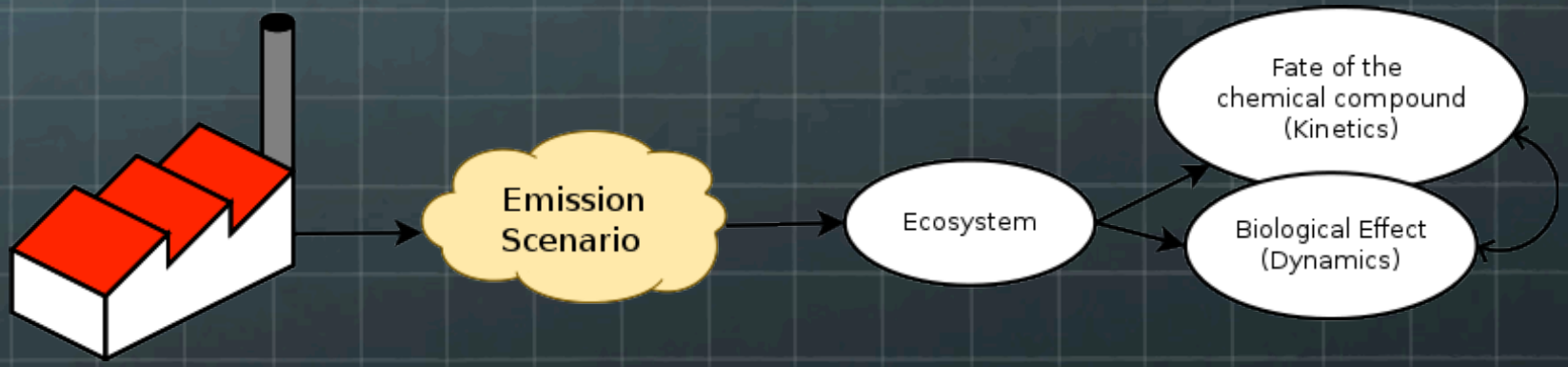
- Common standards for data and model exchange
- Improvement of interoperability
- Reproducibility of QSAR models

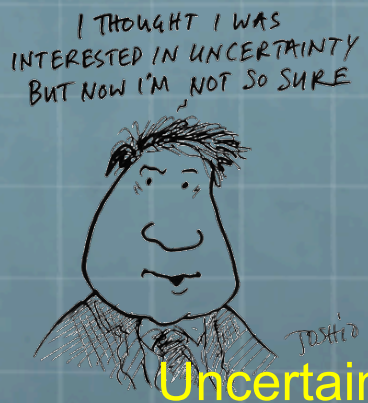
- Computer-Aided Drug Development
- Reduction of the cost for the development of candidates
- Reduction of the number of animal experiments

Problem Statement

- In environmental chemo-informatics the key question we need to answer is *whether the emission of a certain chemical to the ecosystem can cause environmental effects* and, subsequently, to quantify the extent or the likelihood of adverse effects.





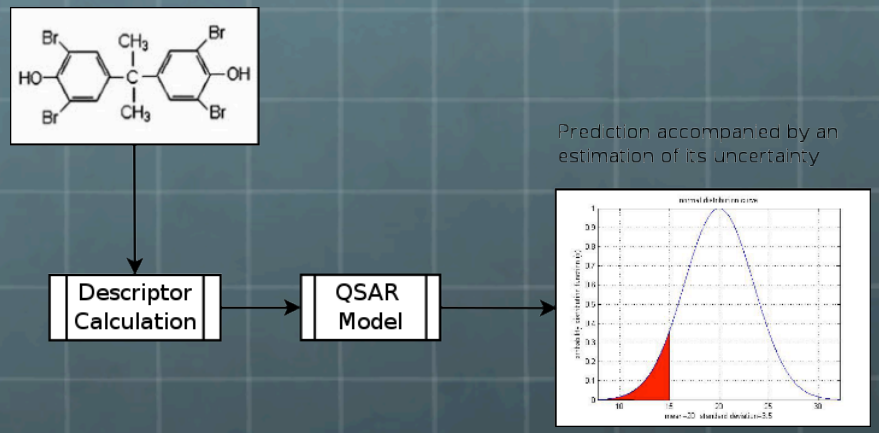


Uncertainty

Uncertainty is ubiquitus in environmental chemo-informatics:

- Uncertainty of the output of QSAR models (modelling error),
- Uncertainty induced by the use of uncertain experimental measurements (including random and systematic errors),
- Mismatch between the modelled ecosystem and the real one including time-varying fluctuations of the local population dynamics and chemical state (pH of water, concentration of ions etc).

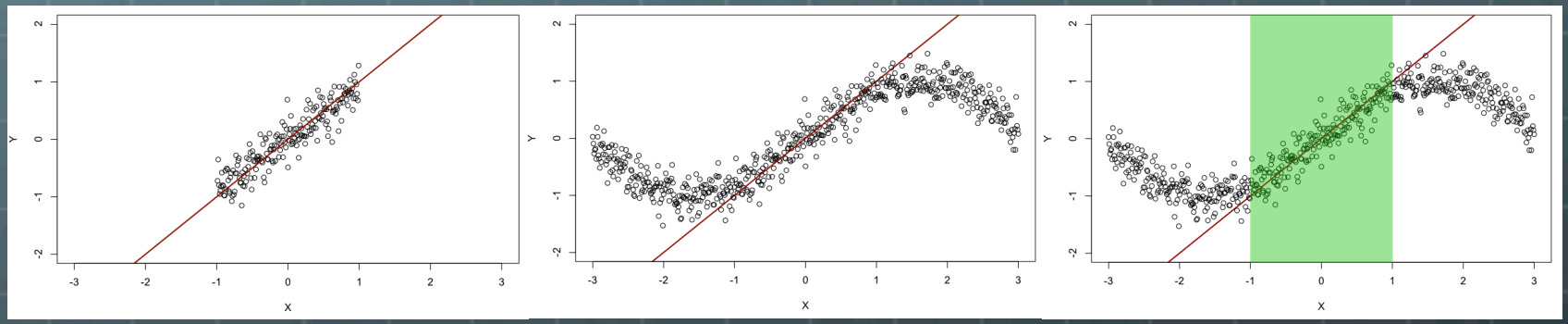
Modelling & Uncertainty

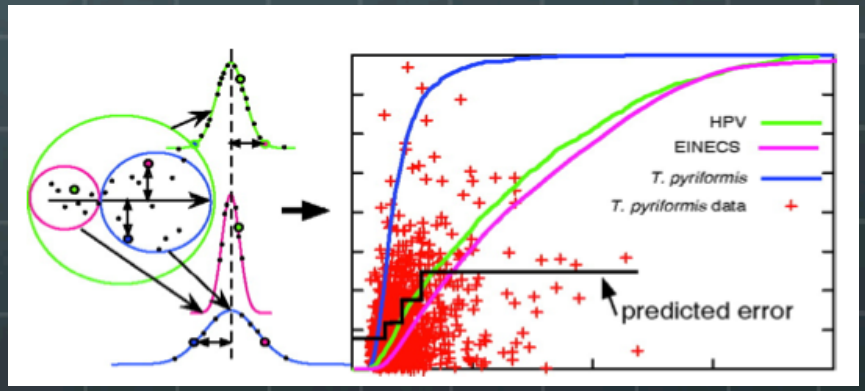


- For arbitrary chemical compounds certain physicochemical parameters may not be known; for these we employ QSAR models.
- We assume that all QSAR predictions are accompanied by an estimation of their uncertainty (in the form of a probability distribution function).

Accuracy of Predictions

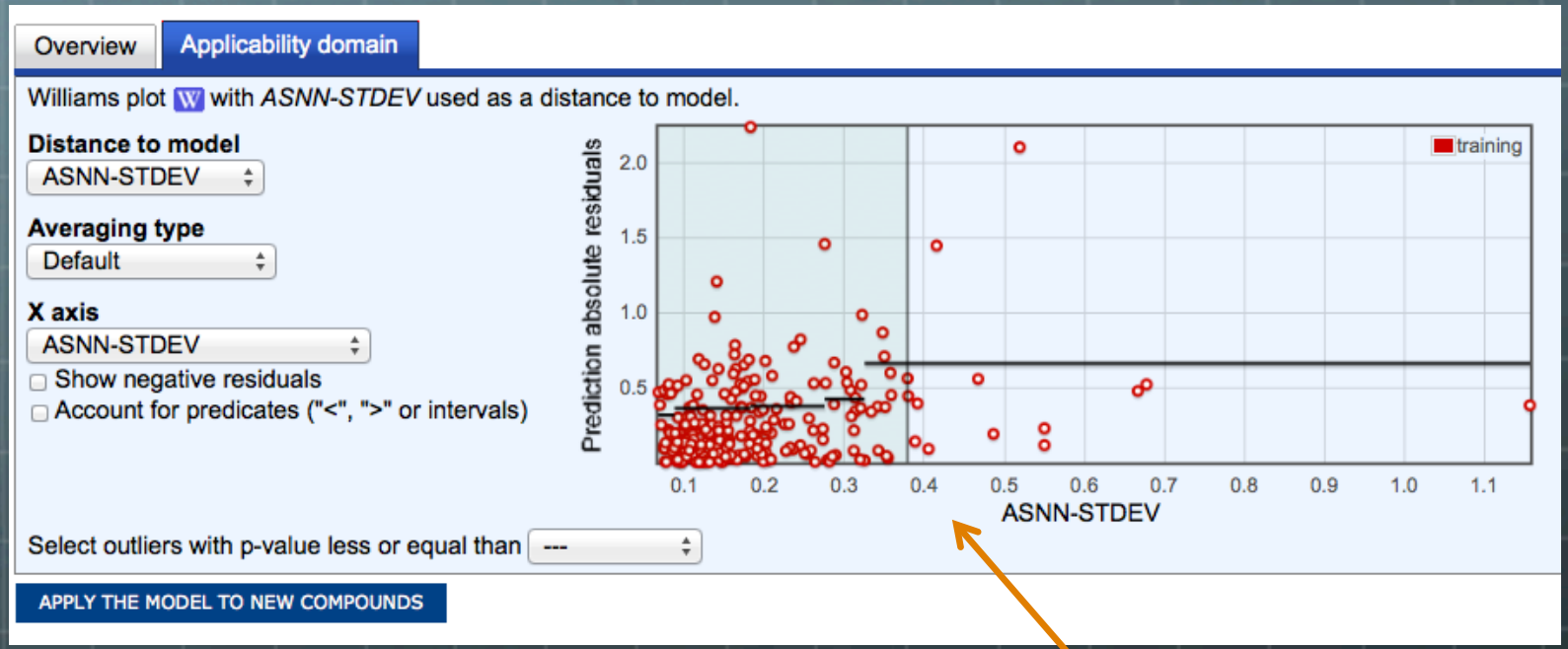
Fact: Models that describe the whole chemical space are still to be found…





I. Tetko, I. Sushko, A. Pandey, H. Zhu, A. Trophsa, E. Papa, T. Oberg, R. Todeschini, D. Fourches and A. Varnek, (2008) "Critical assessment of QSAR Models of environemtal toxicity against Tetrahymena pyriformis: focusing on applicability domain and overfitting by variable selection," J. Chem. Inf. 48(9), 1733-1746.

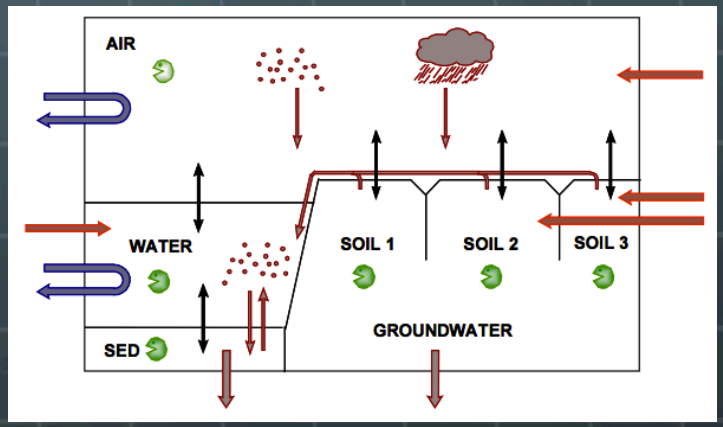
Accuracy Estimation for Regression Models



Abiotic degradation in water

SimpleBox

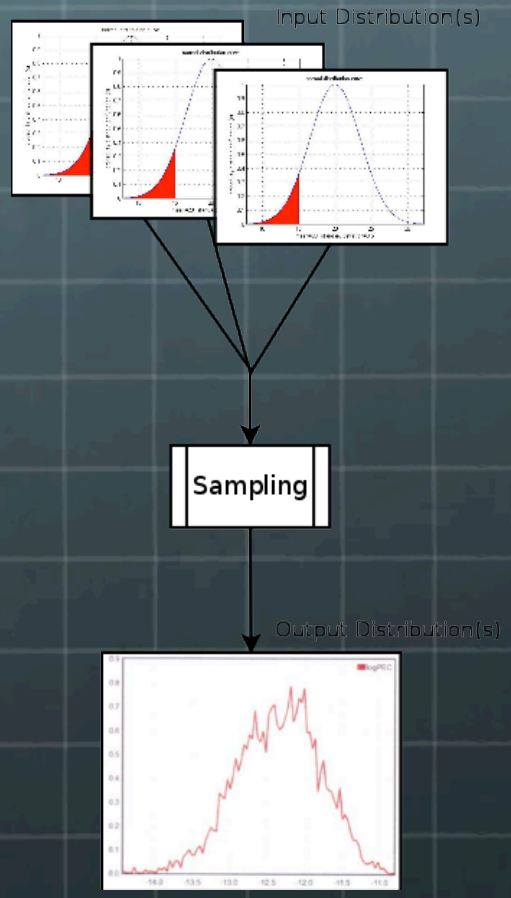
- SimpleBox is an environmental fate model we employed to calculate PEC (Predicted Environmental Concentration)
- SimpleBox assumes the knowledge of a number of physicochemical properties & information in regard to the local ecosystem.

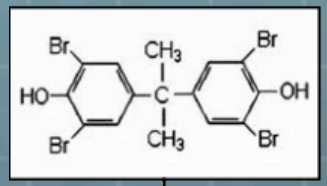


H. A. den Hollander and D. van de Meent (2004). *SimpleBox 3.0: A Multimedia Mass Balance Model for Evaluating the Environmental Fate of Chemicals. RIVM Report 601200003*, 155 pp. Bilthoven, The Netherlands: RIVM, National Institute of Public Health and the Environment.

Monte-Carlo Simulations

- The key question: **How does uncertainty propagate in one's computations?**
- A bit more rigorously: "**How to map a probability distribution function through an (arbitrary) function.**"
- Given that: **In most cases there is no analytical solution available** and this gives rise to **Monte-Carlo simulations**.



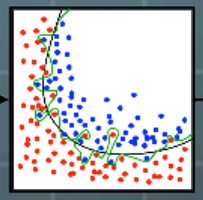


Experimental properties provided by the experimenter (accompanied by an estimation of their uncertainty)

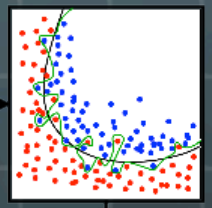
DB with stored physicochemical parameters and their uncertainty

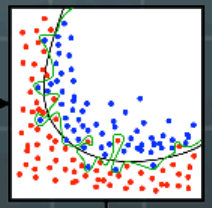
Database of QSPR Thesaurus

Descriptor Calculation



QSAR Predictions accompanied by an estimation of their uncertainty





Simplebox (Estimation of the environmental fate of the chemical compound)

Estimation of the Environmental Fate

H. A. den Hollander and D. van de Meent (2004). *SimpleBox 3.0: A Multimedia Mass Balance Model for Evaluating the Environmental Fate of Chemicals. RIVM Report 601200003*, 155 pp. Bilthoven, The Netherlands: RIVM, National Institute of Public Health and the Environment.

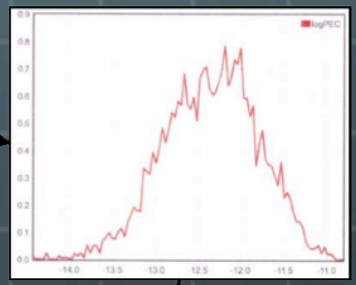
Environmental Risk Assessment

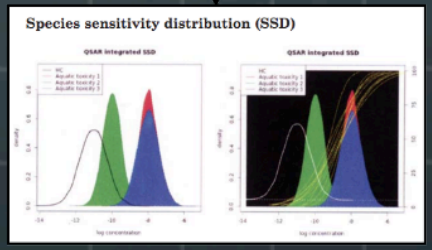
These properties, if available, are provided by the experimenter or are looked up for in the QSPR-Thesaurus database or, otherwise, are predicted by some QSAR model.

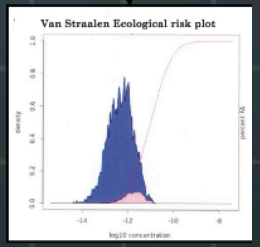
Gas Phase Diffusion Coefficient
Octanol/water partition coefficient
Dissolved phase degradation rate constant at 25°C
Water Phase Diffusion Coefficient
Molecular Weight
Gas/water partition coefficient at 25°C
Vapour Pressure at 25°C
Melting point
Soil Organic Carbon/ Water Partition Coefficien
Junge's constant
Solid/water partition coefficient for standard solids
Standard mass fraction organic carbon in soil/sediment
Concentration of Bacteria in test water
Mineral density sediment and soil
More...

Simplebox

Probabilistic estimation of the environmental fate of a chemical compound using the Simplebox model.





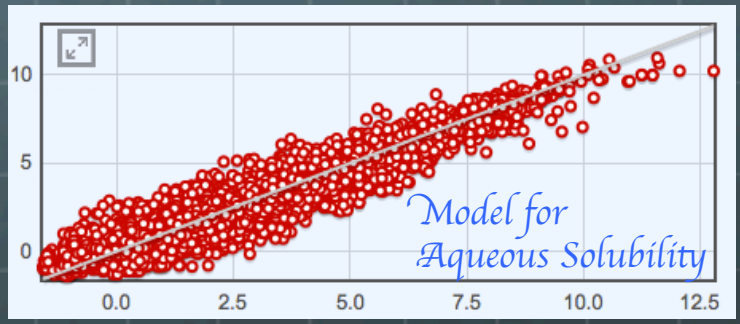


The Species Sensitivity Distribution (SSD) Method is used to estimate the environmental risk.

I. Tetko, P. Sopasakis, P. Kunwar, S. Brandmaier, S. Novotarskyi, L. Charochkina, V. Prokopenko and W. Peijenburg (2013), **Prioritization of Polybrominated Diphenyl Ethers (PBDEs) using the QSPR- Thesaurus Web Tool,** ATLA, 40, pp. 1-9.

Predictive Models

- Machine learning methods used: ANN, OLS.
- All models are validated and are accompanied by an estimation of their domain of applicability *
- For ever prediction there is an estimation of its uncertainty.



The Web Interface

1. Information about the compound

Molecule ID M4475
Molecular Weight 722.48
Name: BDE-183

2. Emission Scenario

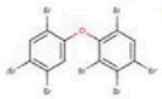
The substance is assumed to be emitted in the air. Please provide the estimated daily emission rate:

Emission	
Emission rate ton/year	10
Emission rate, std.	1

3. Monte-Carlo Iterations

Number 1000

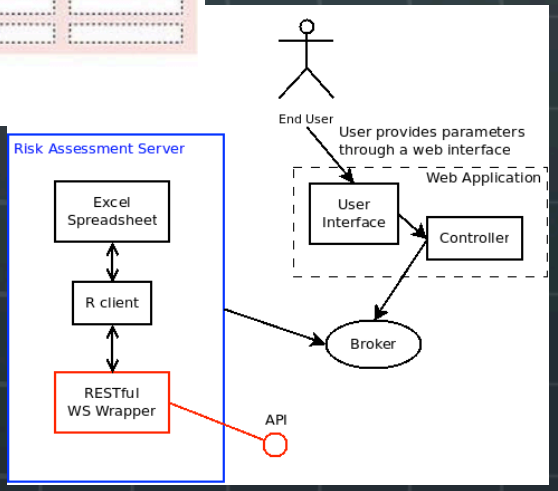
4. Experimental Properties and Uncertainty *(All primary properties are required. The parameters that have empty values are marked as *. The parameters in the advanced options section are optional.)*



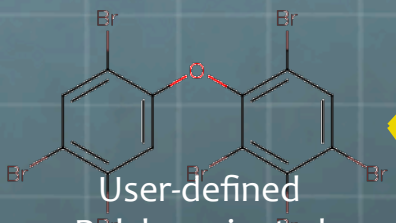
Property	Description	Unit	Use Database Record	Database Record	Use Model	Model	Prediction	Provide Values	Exp. Value	Exp. St. Dev.
MW	Relative Molecular Mass	-	◉	722.48						
Sol25	Water SOLUBILITY at 25 oC	lg([mg.L-1])	○	-2.82log(mg/L)	◉	HMGU: Aqueous Solubility [...]	-3.30±0.700 log(mg/L)	○		
Tm	Melting point	[oC]	○	171.0°C	◉	UI: Melting Point (BFR) [...]	170±20.0 °C (in AD)	○		
Pvap25	VAPOR PRESSURE at 25 oC	lg([Pa])	◉	-6.33log(Pa)	○	UI: Vapor Pressure (BFR) [...]	-5.90±0.200 log(Pa) (in AD)	○		
Koc	Soil Organic Carbon-Water Partitioning Coefficient	lg([-])	○	-	◉	HMGU: Koc [...]	5.20±0.600 log10	○		
kdeg.water	Dissolved phase degradation RATE CONSTANT at 25 oC	lg([s-1])	○	-	◉	HMGU: Abiotic degradation in water [...]	-7.60±0.700 lg(1/s)	○		
kdeg.phot	Photolysis rate	lg([s-1])	○	-5.17-log(s)	◉	HMGU: Photolysis rate [...]	-3.30±0.600 -log(s) (in AD)	○		
Kow	Octanol/water PARTITION COEFFICIENT	lg([-])	◉	8.27log10	○	UI: logKow (PBDE) [...]	8.10±0.200 log10 (in AD)	○		
k0.OHrad	FREQUENCY FACTOR OH radical reaction	lg([cm3.s-1])	○	-	◉	HMGU: Atmospheric OH Rate [...]	-13.0±0.900 log(cm^(3)/(molecule*sec))	○		

Advanced options...

- Screenshot of the QSPR-Thesaurus Web Interface
- See http://qspr-thesaurus.eu



The Web Interface



User-defined Polybrominated Compound

Use Database Record	Database Record
◉	722.48
○	-2.82log(mg/L)

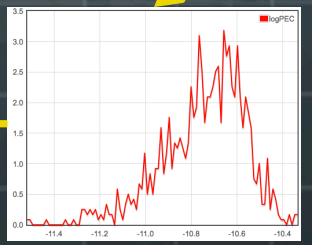
◉ -2 0.1

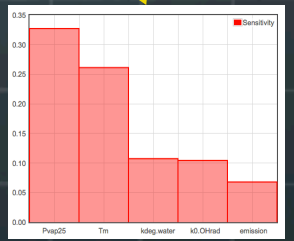
◉ HMGU: Aqueous Solubility [...] -3.30±0.700 log(mg/L) (in AD)

◉ HMGU: Atmospheric OH Rate [...] -13.0±0.900 log(cm^(3)/(molecule*sec)) (**outside AD**)

Inspect the parameters

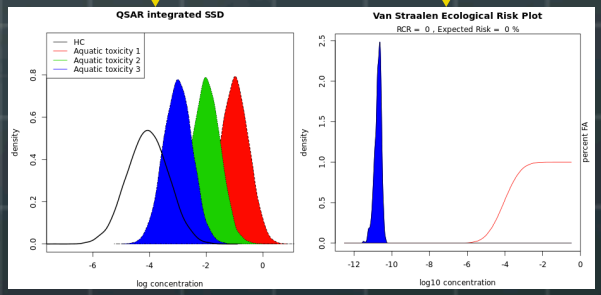
Parameter	Source	Value	Uncertainty
Kow	DatabaseRecord	8.27	0.0
CORG	ExperimentalProperty	0.02	-
RHOsolid	ExperimentalProperty	3.4	-
Pvap25	QSARModel	-5.9	0.2
H0vap	ExperimentalProperty	1.7	-
Sol25	DatabaseRecord	-2.82	0.0
H0sol	ExperimentalProperty	1.0	-
JungeConst	ExperimentalProperty	-0.76	-
Tm	QSARModel	170.0	20.0
kdeg.phot	DatabaseRecord	-5.17	0.0
C.OHrad	ExperimentalProperty	5.7	-
k0.OHrad	ExperimentalProperty	-10.1	0.5
Ea.OHrad	ExperimentalProperty	0.78	-
kdeg.water	QSARModel	-7.6	0.7
BACT.test	ExperimentalProperty	4.6	-
Q.10	ExperimentalProperty	2.0	-
Koc	ExperimentalProperty	5.0	0.0





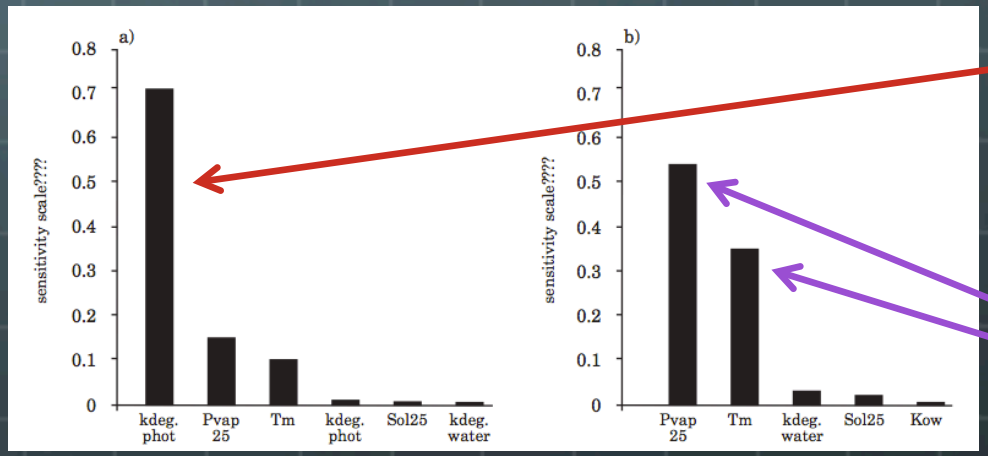
OpenOffice.org File Edit View Insert Format Tools Data Window Help

	A	B	C	D	E
1	sample	HC	Aquatic toxicity 1	Aquatic toxicity 2	Aquatic toxicity 3
2	1	-3.411	-0.363	-1.309	-2.459
3	2	-4.120	-1.645	-2.178	-3.377
4	3	-4.120	-0.209	-1.915	-2.737
5	4	-4.050	-1.012	-1.482	-3.144
6	5	-4.606	-0.752	-2.347	-3.281
7	6	-3.775	-0.799	-1.903	-2.799
8	7	-3.775	-0.944	-1.530	-2.927
9	8	-4.719	-0.821	-2.070	-3.492
10	9	-3.293	-0.659	-1.751	-2.387



Sensitivity analysis

- Sensitivity analysis reveals the extent on which the outcome of the environmental risk assessment depends on the various physicochemical and/or biological parameters considered in the study
- It, thus, provides evidence for a mechanistic interpretation of the key factors that create the environmental hazard.



In this study (on PBDE-177) it was found that the photo-degradation rate constant was the most determining factor.

If photodegradation is not taken into account, then the vapour preseure at 25C and the melting point become the most influential parameters.

I. Tetko, P. Sopasakis, P. Kunwar, S. Brandmaier, S. Novotarskyi, L. Charochkina, V. Prokopenko and W. Peijenburg (2013), **Prioritization of Poly-brominated Diphenyl Ethers (PBDEs) using the QSPR- Thesaurus Web Tool**, ATLA, 40, pp. 1-9.

Recapitulation

- We have developed a *web tool* for the exemplification of the use of QSAR models for the *fate*, *effect* and *risk* assessments of chemical compounds.
- The fate assessment was carried out in terms of the *Predicted Environmental Concentration (PEC)*. The *effect* was quantified by the *Predicted No-Effect Concentration (PNEC)*.
- SimpleBox and SSD (Species Sensitivity Distribution) were used to model the fate and the effect of PBDEs on the ecosystem.

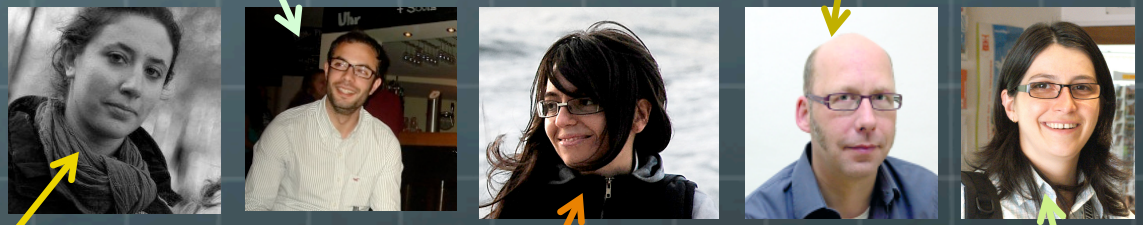
Acknowledgements

This work was greatly supported by the following people we wish to thank:

- Dr. Ullirika Sahlin (Linnaeus University),
- Prof. Mark Huijbregts (Radboud University) and
- Mr. Sarfaz Iqbal (Linnaeus University).

This study was partly supported by the EU through the CADASTER project (FP7-ENV-2007- 212668) and the FP7 MC ITN project Environmental Chemoinformatics (ECO), grant agreement No. 238701, and the GO-Bio 1B BMBF project iPRIOR, grant agreement No. 315647.

Acknowledgements



Matteo Cassotti

Kai Zillessen

Alessandra Pirovano

Valentina Zingarelli

Ioana Oprisiu