Electronic Supplementary Information

Eawag-Soil in enviPath: A new resource for exploring regulatory pesticide soil biodegradation pathways and half-life data

Diogo A. R. S. Latino[a,*], Jörg Wicker[b], Martin Gütlein[b], Emanuel Schmid[c], Stefan Kramer[b], Kathrin Fenner[a,d]

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

[b] Institute of Computer Science, Johannes Gutenberg University Mainz, 55128 Mainz, Germany

[c] Scientific IT Services, ETH Zürich, 8092 Zürich, Switzerland

[d] Department of Chemistry, University of Zürich, 8057 Zürich, Switzerland

*Corresponding author: diogo.latino@eawag.ch, Tel.: +41 58 765 5737

ESI Table of Contents

S1 *Eawag-Soil* metadata and conventions

***Scenario name*:** The *scenario name* is generated using the author name and year of the study, "Author name, Year" and in some cases also the study name, "Study name, Author name, Year".

***Scenario description*:** Small description of the scenario with information concerning the date of the submission of the Draft Assessment Report (DAR) (date of the dossier on the first page).

***Soil origin*:** Information on the soil source, typically indicating the place where the sample was collected - city, canton or state and country.

***Soil description*:** Information on *soil texture classification* and *% sand*, *% clay* and % *silt*. The *soil texture* is a qualitative classification used for determination of soil classes based on their physical texture.

***Soil texture classification system*:** Information on the *soil texture classification system* used to derive the *soil texture* given in *soil description*. Information from the USDA classification system is preferred if information from more than one classification system is available.

***Redox conditions*:** All data stored in the package *Eawag-Soil* are from aerobic studies. Data from anaerobic-aerobic and anaerobic studies are not considered.

***Acidity, pH*:** pH at which the study was conducted. For studies where pH was determined in different soil solutions (i.e., H_2O, KCl, $CaCl_2$), only one value is stored using the following order of preference: $CaCl_2$ > KCl > H_2O. If available, the pH range of the experiment is stored.

Temperature: Temperature at which the experiment was performed in °C.

***Water storage capacity*:** Information on how much water the soil can hold under a given "condition", in g/100g soil or mass-%. The "condition" indicates under which conditions the water storage capacity was measured (i.e., "(maximum) water holding capacity", pF 2, 1/3 bar etc.).

***Experimental humidity*:** Indicates at what percentage of the *water storage capacity* the experiment was actually performed.

***Organic content, OC*:** Information on *soil organic matter (OM)* or *organic carbon (OC)* content in mass-%. If both values were available in the DAR, the value that has actually been experimentally measured (and not derived from the respective other value) is stored.

***Cation exchange capacity, CEC*:** Information on the number of exchangeable cations that a soil is capable of holding per dry weight at a given pH value and available for exchange with the soil water solution, measured in milliequivalent of hydrogen per 100g of dry soil.

***Bulk density*:** Weight of dry soil for a given soil volume in g/cm^3.

***Biomass*:** Biomass is the part of *soil organic matter*, or *organic carbon*, that is made up by living organisms, mostly bacteria and fungi. Data is stored in µg of carbon/g soil.

***Spike concentration/Spike compound*:** The molecular formula of the *spike compound*, including isotopic information, using SMILES strings. For indicating the specific position of the ^{14}C-label, the following rules were followed: *i)* if a specific position is indicated in the structures drawn in the DAR, the structure is stored with the label at that specific position; *ii)* If the radioactivity "mark" (in general, star symbol) is in the middle of the aromatic ring or if the compound is given as [U^{14}C]-compound, then the structure is stored with the ^{14}C-label attached to all C atoms in the aromatic ring.

***DT50* values:** The DT50 value (50% dissipation time) is the time required for 50% of the initial concentration of the compound under study to dissipate (in days). Only DT50 values measured at the original temperature of the experiment are considered. While the definition of DT50 is independent of a specific kinetic model, in practice DT50 values are calculated by assuming specific kinetic models. When several DT50 values are given for the same set of data (e.g., original submission, re-calculated by a "reviewing committee", different kinetic models etc.), only the value with the highest quality is stored (e.g., based on quality of fit, R^2 values, comments in the text etc.). A comment about whether or not the DT50 values is derived from a first-order kinetic model and a comment about the reason for selecting that specific DT50 value (if needed) are also stored.

There are a number of situations that could lead to different DT50s for the same compound and experimental conditions, which had to be resolved during data curation. The following conventions explain how the DT50s were treated and saved in each case:

- Same compound labelled in more than one position:
  Both DT50s were saved in *Eawag-Soil*.
- Experiments with different durations:
  In some cases, DT50s are available from two studies with the same experimental conditions where the only difference is the duration of the study, e.g., 120 days and more than 120 days. In this case, the DT50 from the experiment performed in 120 days is saved and DT50s from experiments with longer duration are not considered.
- Spike compounds are enantiomers of each other:
  Both DT50s were saved in *Eawag-Soil*.
- DT50s normalized to specific experimental conditions, e.g., *water storage capacity conditions* or *temperature*:
  The non-normalised DT50 value were saved if available. If not available the normalised value was saved with a comment.
- DT50s derived from SFO models that also considered metabolite formation and degradation:
  If available only the DT50 from the "parent only model" were considered.
- DT50s were back-calculated from DT90:

DT50s back-calculated from DT90 were not considered.

Table S1. Statistical measures for *DT50* and experimental parameters *% sand*, *% silt*, *% clay*, *pH*, *temperature*, *water storage capacity*, *% humidity*, *organic content* (*OC)*, *cation exchange capacity* (*CEC)*, *bulk density*, *biomass start*, *biomass end* and *spike concentration*.

Parameter	**Dimension**	**N**		**Mean**	**Median**	**Mode**	**Std. Deviation**	**Skewness**	**Kurtosis**	**Range**	**Minimum**	**Maximum**
		Valid	**Missing**									
DT50	days	4716	0	88.8	24.0	1.00	207	7.07	77.2	3689	0.003	3690
% Sand	–	4108	608	51.9	55.0	51.0	24.6	-0.29	-1.07	99.0	0.00	99.0
% Silt	–	4114	602	32.0	28.0	24.0	18.6	0.64	-0.407	88.6	0.00	88.6
% Clay	–	4132	584	16.0	12.6	12.0	11.0	1.49	3.69	94.2	0.10	94.3
Acidity, pH	–	4641	75	6.62	6.70	7.30	0.888	-0.37	-0.510	5.20	3.60	8.80
Temperature	°C	4659	57	20.1	20.0	20.0	3.50	-0.68	7.41	48.0	1.00	49.0
Water Storage Capacity	g water / 100 g dry soil	3998	718	40.2	38.9	29.8	18.1	0.60	0.863	135.06	1.54	136
% Humidity	–	4350	366	55.4	45.0	40.0	21.6	1.06	-0.142	95	5.00	100
Organic Content (OC)	g OC/100 g soil	4540	176	1.81	1.62	1.30	1.08	2.02	7.21	9.98	0.02	10.0
Cation Exchange Capacity (CEC)	mEq/100 g soil	3785	931	14.0	12.2	10.0	7.92	1.66	4.25	58.8	1.20	60.0
Bulk Density	g/cm^3	1741	2975	1.34	1.40	1.50	0.282	0.99	5.52	2.66	0.001	2.66
Biomass Start	µg C/g	3147	1569	408	312	250	361	2.26	7.14	2444	1.00	2445
Biomass End	µg C/g	2651	2065	345	257	200	339	2.83	10.9	2451	0.05	2452
Spike Concentration	mg/kg dry soil	3917	799	1.28	0.400	0.20	2.55	4.42	25.8	25.0	0.003	25.0

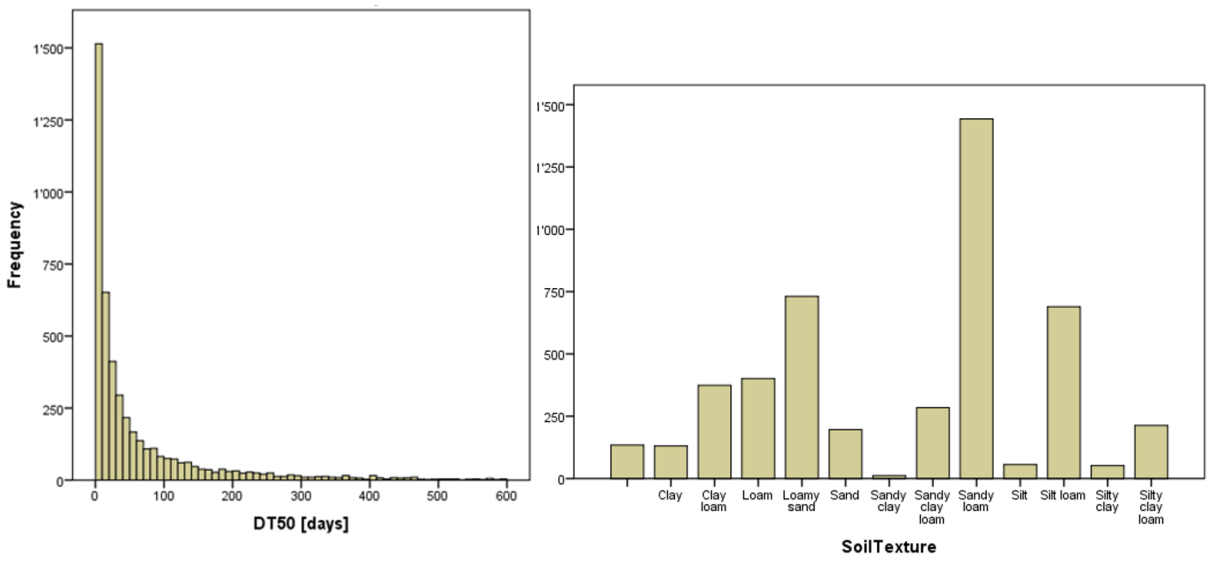


Figure S1. Histograms for DT50 and soil texture.

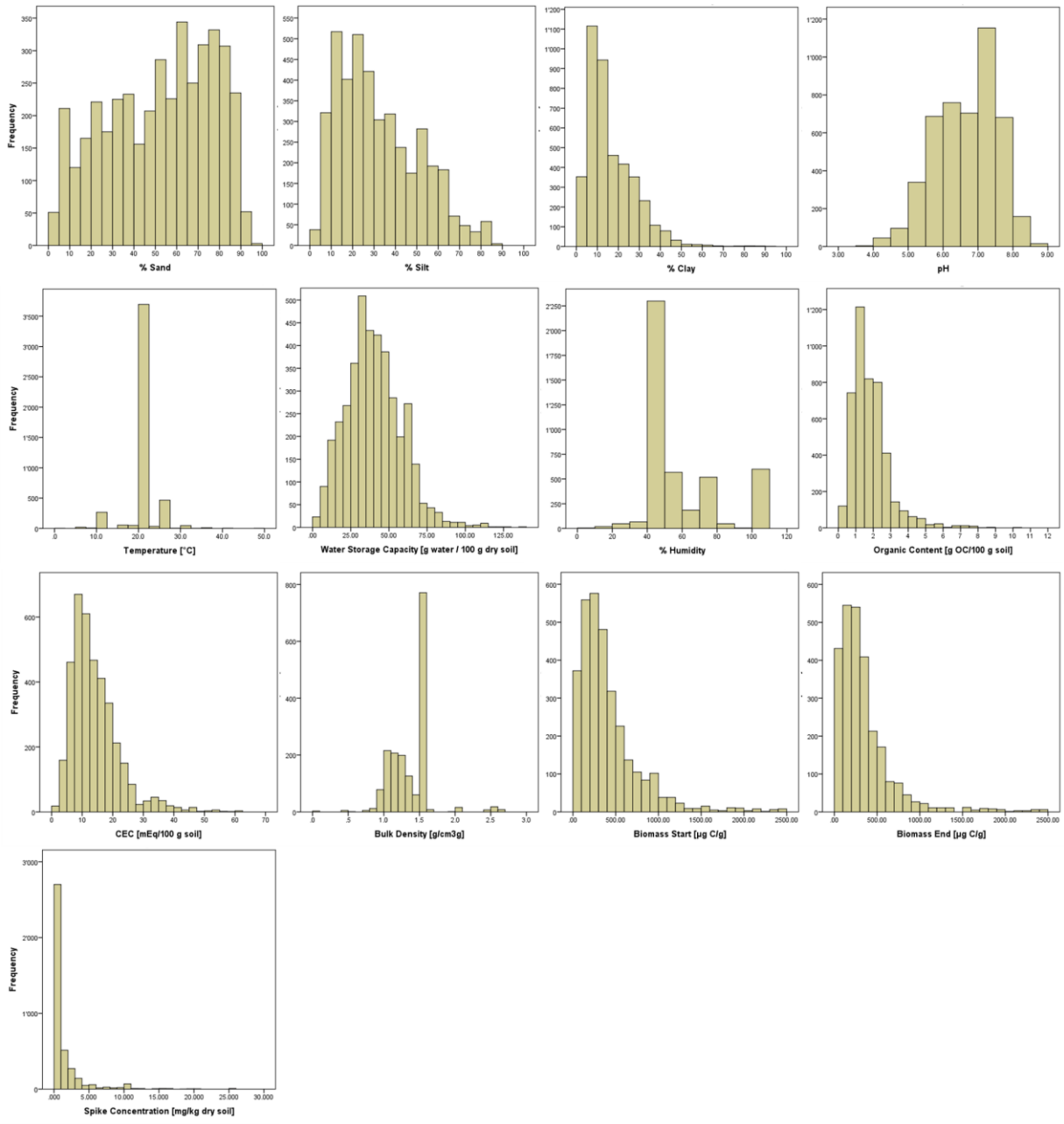


Figure S2. Histograms for all numerical experimental parameters.

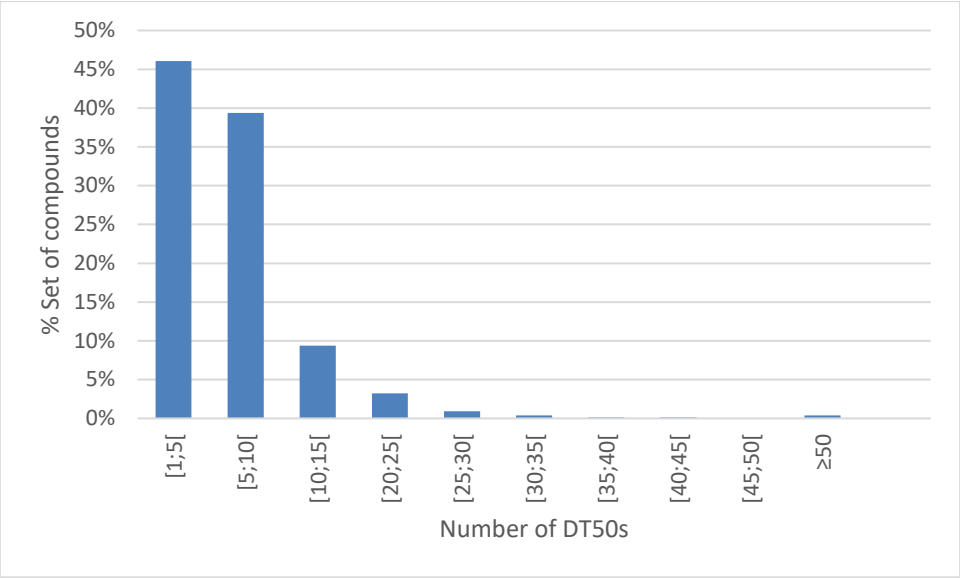


Figure S3. Histogram showing the number of DT50 values available per compound.

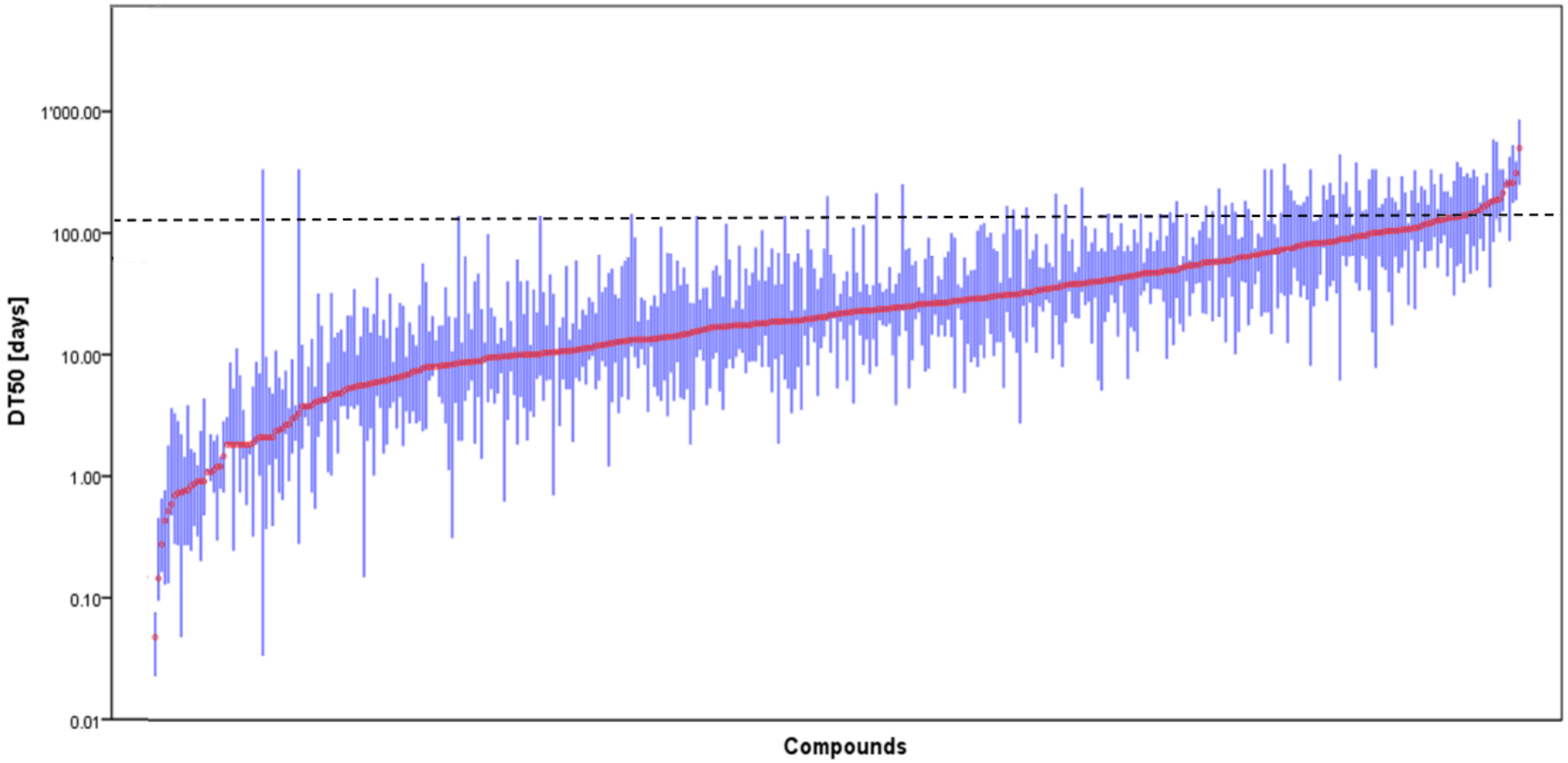


Figure S4. Median DT50 values (red diamonds) and DT50 distributions (minimum to maximum) for 777 compounds with associated DT50 values in *Eawag-Soil*. The dashed line indicates a persistence criterion in soil of 120 days.

Table S2. Compounds with at least 10 DT50s values and corresponding SMILES string, minimum, maximum, range and median DT50.

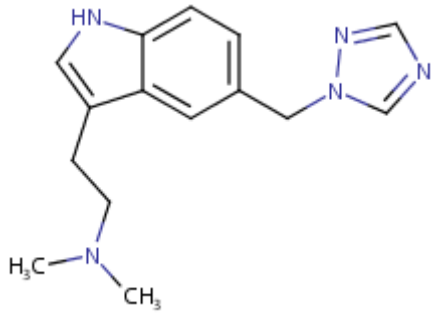
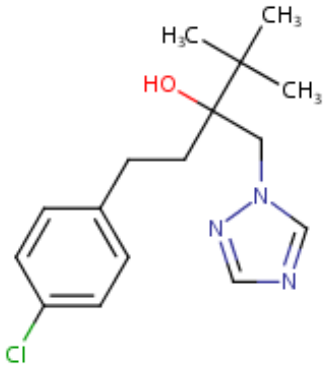
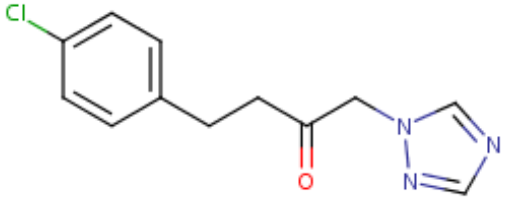
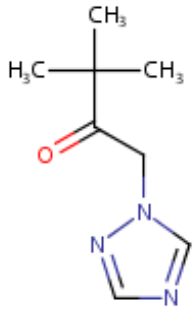
Compound	**SMILES**	**Minimum DT50**	**Maximum DT50**	**Median DT50**	**DT50 Range**	**N. DT50s**
Trinexapac-ethyl	CCOC(=O)C1CC(=O)/C(=C(\C2CC2)/O)/C(=O)C1	0.03	0.32	0.15	0.29	10
Acibenzolar-S-Methyl	CSC(=O)C1=CC=CC2=C1SN=N2	0.03	0.98	0.19	0.95	17
Pinoxaden	CCC1=CC(=CC(=C1C2=C(N3CCOCCN3C2=O)OC(=O)C(C)(C)C)CC)C	0.08	1.81	0.30	1.73	16
Cycloxydim	CCC/C(=N\OCC)/C1=C(CC(CC1=O)C2CCCSC2)O	0.08	2.70	0.32	2.62	10
3-Chloroallyl alcohol	C(=C\Cl)/CO	0.10	0.60	0.40	0.50	10
Fluroxypyr	CCCCCCC(C)OC(=O)COC1=NC(=C(C(=C1Cl)N)Cl)F	0.30	1.80	0.75	1.50	12
Fosetyl-AL	CCO[P+](=O)[O-].CCO[P+](=O)[O-].CCO[P+](=O)[O-].[Al+3]	1.00	2.00	1.00	1.00	13
Quizalofop-P-tefuryl	CC(C(=O)OCC1CCCO1)OC2=CC=C(C=C2)OC3=NC4=CC=C(C=C4N=C3)Cl	1.00	7.87	1.00	6.87	12
Cymoxanil	CCNC(=O)NC(=O)/C(=N/OC)/C#N	0.10	4.30	1.05	4.20	10
1-naphthoic acid (M2)	C1=CC2=C(C=C1)C(=CC=C2)C(=O)O	0.46	6.00	1.20	5.54	12
Ester sulfonamide (IN-D5803)	COC(=O)C1=C(C=CC=C1)S(=O)(=O)N	0.90	12.3	2.60	11.4	14
2-(1-naphtyl)acetic acid (M1)	C1=CC2=C(C=C1)C(=CC=C2)CC(=O)O	1.22	44.4	2.96	43.2	14
Florasulam	COC1=NC=C(C2=NC(=NN21)S(=O)(=O)NC3=C(C=CC=C3F)F)F	0.46	45.0	3.50	44.5	17
Pyroxsulam (XDE-742)	COC1=NC2=NC(=NN2C(=C1)OC)NS(=O)(=O)C3=C(N=CC=C3C(F)(F)F)OC	0.80	16.7	3.60	15.9	25
Tralkoxydim	CCC(=NOCC)C1C(=O)CC(CC1=O)C2=C(C)C=C(C)C=C2C	1.10	31.0	4.50	29.9	10
MITC	CN=C=S	0.97	32.7	5.62	31.7	13
Amitrole	C1=NNC(=N1)N	1.40	94.5	6.80	93.1	11
2,4-D	CC1=CC=C(C(=C1)C)OCC(=O)O	1.48	59.3	7.00	57.8	38
Dimethachlor	CC1=CC=CC(=C1N(CCOC)C(=O)CCl)C	3.31	19.8	7.15	16.5	12
Pyriproxyfen	CC(COC1=CC=C(C=C1)OC2=CC=CC=C2)OC3=NC=CC=C3	2.80	20.0	7.30	17.2	10
Prohexadione-calcium	CCC(=O)C1=C(CC(CC1=O)C(=O)[O-])[O-].[Ca+2]	1.10	279	7.80	278	12
Glufosinate ammonium (GA)	CP(=O)(CCC(C(=O)O)N)[O-].[NH4+]	5.00	18.0	8.00	13.0	13
Glyphosate	C(C(=O)O)NCP(=O)(O)O	1.02	60.2	8.07	59.2	11
Metalaxyl-M	CC1=CC=CC(=C1N([C@H](C)C(=O)OC)C(=O)COC)C	3.30	73.1	8.13	69.8	11
3-Chloroacrylic acid	C(=C\Cl)/C(=O)[O-]	0.70	30.0	8.20	29.3	10
S13 (2-amino-4,6-dimethoxypyrimidine)	COC1=NC(=NC(=C1)OC)N	2.90	197	8.86	194	16
Methomyl	C/C(=N\OC(=O)NC)/SC	4.00	31.0	8.95	27.0	10
Etofenprox	CCOC1=CC=C(C=C1)C(C)(C)COCC2=CC=CC(=C2)OC3=CC=CC=C3	2.80	25.0	9.00	22.2	12
Phosphine	P	0.24	13.6	9.15	13.4	12
Acid sulphonamide (IN-D5119)	C1=CC(=C(C=C1)S(=O)(=O)N)C(=O)O	1.00	116	9.50	115	18
IN-V7160	CC1=NC(=NC(=N1)NC(=O)N)OC	3.30	60.0	9.55	56.7	14

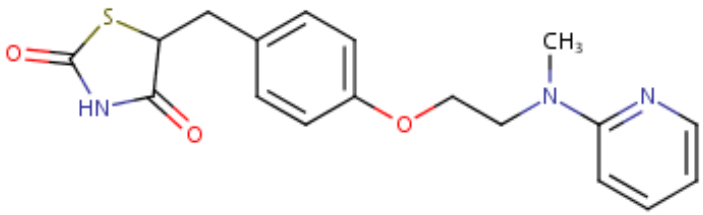
Trinexapac-acid	C1CC1/C(=C/2\C(=O)CC(CC2=O)C(=O)O)/O	2.50	26.5	9.55	24.0	10
NOA 407854 (M2)	CCC1=CC(=CC(=C1C2C(=O)N3CCOCCN3C2=O)CC)C	1.11	61.3	9.62	60.2	16
Tribenuron-methyl	CC1=NC(=NC(=N1)N(C)C(=O)NS(=O)(=O)C2=C(C=CC=C2)C(=O)OC)OC	3.00	46.0	10.0	43.0	14
Acetochlor	CCC1=C(C(=CC=C1)C)N(COCC)C(=O)CCl	0.28	44.0	10.2	43.7	50
BH 517-TSO2 E/Z-isomer	CCC/C(=N\OCC)/C1=C(CC(CC1=O)C2CCCS(=O)(=O)C2)O	5.10	24.4	10.5	19.3	13
Isoproturon	CC(C)C1=CC=C(C=C1)NC(=O)N(C)C	7.08	29.6	11.1	22.5	14
DE-535 acid	CC(C(=O)O)OC1=CC=C(C=C1)OC2=NC=C(C=C2Cl)C(F)(F)F	4.00	27.0	12.0	23.0	18
Flupyrsulfuron-methyl	COC1=NC(=NC(=C1)OC)NC(=O)NS(=O)(=O)C2=NC(=CC=C2C(=O)OC)C(F)(F)F	7.10	51.9	12.3	44.8	11
Fluroxypyr ester	CCCCCCC(C)OC(=O)COC1=NC(=C(C(=C1Cl)N)Cl)F.C(C(=O)O)OC1=NC(=C(C(=C1Cl)N)Cl)F	2.90	82.5	12.8	79.6	12
Desmethyl-isoproturon	CC(C)C1=CC=C(C=C1)NC(=O)NC	2.20	39.5	13.3	37.3	12
O-desmethyl triazine amine (IN-B5528)	CC1=NC(=NC(=N1)N)O	3.30	88.6	13.4	85.3	12
Metazachlor	CC1=CC=CC(=C1N(CN2C=CC=N2)C(=O)CCl)C	3.10	109	13.6	106	49
Fenpropimorph	CC(CC1=CC=C(C=C1)C(C)(C)C)CN2C[C@H](C)O[C@H](C)C2	9.00	181	14.0	172	11
Saccharin (IN-00581)	C1=CC2=C(C=C1)S(=O)(=O)NC2=O	3.00	237	14.7	234	21
Phenylurea (IN-B5685)	COC(=O)C1=C(C=CC=C1)S(=O)(=O)NC(=O)N	5.00	45.0	14.8	40.0	12
MB1 CGA359009	CC1=NNC(=O)N(C1O)/N=C/C2=CC=CN=C2	2.04	123	15.0	121	10
Carbofuran	CC1(C)CC2=CC=CC(=C2O1)OC(=O)NC	6.10	110	15.0	104	10
Pyridinol (DCP)	C1(=C(F)N=C(C(=C1N)Cl)O)Cl	3.10	85.2	16.0	82.1	12
Hexythiazox	CC1C(C2=CC=C(C=C2)Cl)SC(=O)N1C(=O)NC3CCCCC3	1.00	35.2	16.5	34.2	10
CGA_50266	CC1=CC=CC(=C1N(CCOC)C(=O)C(=O)O)C	6.80	254	19.3	248	10
Hydroxy Propaquizafop acid (CGA_294972)	CC(C(=O)O)OC1=CC=C(C=C1)OC2=NC3=CC=C(C=C3N=C2O)Cl	7.00	217	21.0	210	17
5-OH Florasulam	C1=CC(=C(C(=C1)F)NS(=O)(=O)C2=NN3C(=N2)C(=CN=C3O)F)F	6.99	86.0	21.4	79.0	14
Sulcotrione	CS(=O)(=O)C1=CC=C(C(=C1)Cl)C(=O)C2C(=O)CCCC2=O	3.70	137	21.9	133	12
Metalaxyl CGA48988	CC1=CC=CC(=C1N(C(C)C(=O)OC)C(=O)COC)C	8.58	55.2	21.9	46.6	11
MB5 CGA363431	CC1=NNC(=O)N(C1O)/N=C/C2=CNC(=O)C=C2	1.02	123	22.0	122	10
Bentazone	CC(C)N1C(=O)C2=C(C=CC=C2)NS1(=O)=O	9.60	56.0	22.1	46.4	16
Propamocarb hydrochloride	CCCOC(=O)NCCCN(C)C.Cl	5.12	267	22.1	262	22
CGA210007	C1=CC2=C(C(=C1)C(=O)O)SN=N2	4.09	108	22.1	104	16
Amidosulfuron	CN(S(=O)(=O)C)S(=O)(=O)NC(=O)NC1=NC(=CC(=N1)OC)OC	2.20	29.0	22.4	26.8	10
Propaquizafop acid (CGA_287422)	CC(C(=O)O)OC1=CC=C(C=C1)OC2=NC3=CC=C(C=C3N=C2)Cl	7.00	119	22.5	112	15
O-desmethyl metsulfuron-methyl (IN-B5067)	CC1=NC(=NC(=N1)NC(=O)NS(=O)(=O)C2=C(C=CC=C2)C(=O)OC)O	2.40	85.3	22.6	82.9	17
Lufenuron (CGA_184699)	C1=CC(=C(C(=C1)F)C(=O)NC(=O)NC2=CC(=C(C=C2Cl)OC(C(C(F)(F)F)F)(F)F)Cl)F	13.0	83.1	23.7	70.1	11
Metamitron	CC1=NN=C(C2=CC=CC=C2)C(=O)N1N	3.40	49.5	24.4	46.1	28
Avermectin B1a	CC[C@H](C)[C@@H]1[C@@H](C)C=C[C@]2(C[C@@H]3C[C@@H](C/C=C(\C)/[C@H]([C@@H](C)/C=C/C=C/4\CO[C@@H]5[C@@H](C(=C[C@@H](C(=O)O3)[C@]45O)C)O)O[C@H]6C[C@@H]([C	11.2	65.7	24.4	54.5	15

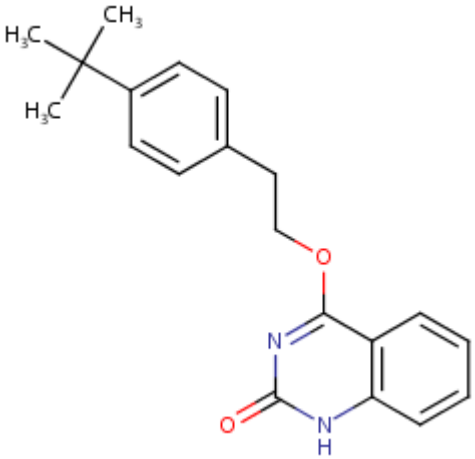
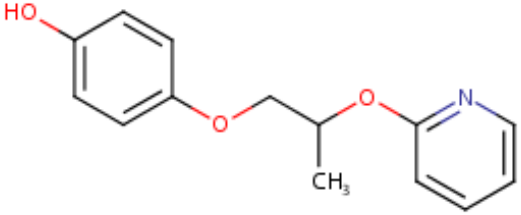
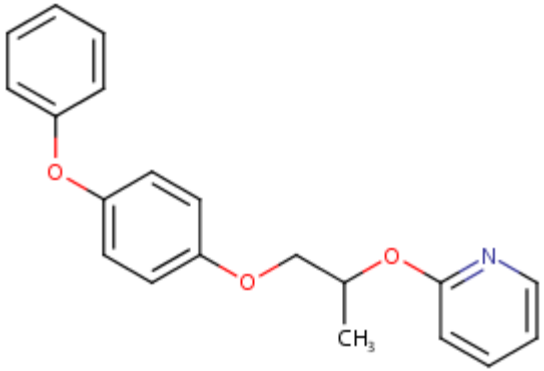
	@H]([C@H](C)O6)O[C@H]7C[C@@H]([C@H]([C@H](C)O7)O)OC)OC)O2)O1					
Chlorsulfuron	CC1=NC(=NC(=N1)NC(=O)NS(=O)(=O)C2=C(C=CC=C2)Cl)OC	2.80	232	28.1	229	51
Pyridafol (CL-9673)	C1=CC=C(C=C1)C2=NN=C(C=C2O)Cl	15.5	46.6	28.5	31.1	10
Lambda-Cythalothrin	CC1(C)C(/C=C(/C(F)(F)F)\Cl)C1C(=O)OC(C#N)C2=CC=CC(=C2)OC3=CC=CC=C3	7.40	248	28.8	241	10
Cyromazine	C1CC1NC2=NC(=NC(=N2)N)N	2.70	56.0	31.1	53.3	12
Beta-cypermethrin	CC1(C)C(/C=C(/Cl)\Cl)C1C(=O)OC(C#N)C2=CC=CC(=C2)OC3=CC=CC=C3	15.2	73.7	31.6	58.5	10
combined Propaquizafop & Propaquizafop_acid	C/C(=N/OCCOC(=O)C(C)OC1=CC=C(C=C1)OC2=NC3=CC=C(C=C3N=C2)Cl)/C.CC(C(=O)O)OC1=CC=C(C=C1)OC2=NC3=CC=C(C=C3N=C2)Cl	4.10	139	32.0	135	11
2-naphthyloxyacetic acid	C1=CC2=C(C=C1)C=C(C=C2)OCC(=O)O	4.10	180	37.0	176	22
M2 CGA294849	CC1=NNC(=O)N(C1=O)N	3.65	109	37.6	105	11
Metsulfuron acid (IN_F5438)	CC1=NC(=NC(=N1)NC(=O)NS(=O)(=O)C2=C(C=CC=C2)C(=O)O)OC	7.90	79.1	38.8	71.2	21
Metsulfuron-methyl	CC1=NC(=NC(=N1)NC(=O)NS(=O)(=O)C2=C(C=CC=C2)C(=O)OC)OC	9.00	80.5	39.0	71.5	22
Clopyralid	C1=C(C(=NC(=C1)Cl)C(=O)O)Cl	7.00	215	39.0	208	18
Quinmerac	CC1=CC2=C(C(=C(C=C2)Cl)C(=O)O)N=C1	16.0	204	41.0	188	27
Cyprodinil (CGA 219417)	CC1=NC(=NC(=C1)C2CC2)NC3=CC=CC=C3	13.0	113	42.0	100	12
NOA409045	CC1=CC=CC(=C1N([C@H](C)C(=O)O)C(=O)COC)C	4.15	200	42.4	196	10
Sulfosulfuron	CCS(=O)(=O)C1=C(N2C=CC=CC2=N1)S(=O)(=O)NC(=O)NC3=NC(=CC(=N3)OC)OC	14.5	365	43.0	350	12
BH 518-2	C1=CC(=C(C2=NC=C(C=C12)C(=O)O)C(=O)O)Cl	12.9	385	45.5	372	18
Fenamiphos-sulfoxide	CCOP(=O)(NC(C)C)OC1=CC=C(C(=C1)C)S(=O)C	10.0	198	48.0	188	19
NOA 447204 (M3)	CCC1=CC(=CC(=C1C2(C(=O)N3CCOCCN3C2=O)O)CC)C	25.7	541	51.3	516	11
IN-JV460	COC1=NC(=NC(=C1)OC)C2C3=C(C=C(C(F)(F)F)N=C3)C(=O)NC2=O	46.1	630	57.0	584	11
Flupyradifurone	C1=C(C=NC(=C1)Cl)CN(CC(F)F)C2=CC(=O)OC2	33.0	242	57.3	209	16
Mandipropamid	C#CCOC1=CC=C(CCNC(=O)C(C2=CC=C(C=C2)Cl)OCC#C)C=C1OC	12.6	159	59.4	146	23
CGA62826	CC1=CC=CC(=C1N(C(C)C(=O)O)C(=O)COC)C	3.88	138	60.4	134	11
MT1, GS 26379	CC(C)(C)NC1=NC(=NC(=N1)Cl)N	21.7	113	61.8	91.1	19
Spiroxamine	CCCN(CC)CC1COC2(CCC(CC2)C(C)(C)C)O1	35.0	133	63.9	98.1	12
Esfenvalerate	CC(C)[C@@H](C1=CC=C(C=C1)Cl)C(=O)O[C@H](C#N)C2=CC=CC(=C2)OC3=CC=CC=C3	27.1	115	66.5	87.8	12
Aclonifen	C1=CC=C(C=C1)OC2=CC=C(C(=C2Cl)N)[N+](=O)[O-]	34.0	207	74.5	173	10
Mandestrobin	CC1=CC=C(C)C(=C1)OCC2=C(C=CC=C2)C(C(=O)NC)OC	49.0	288	75.7	239	10
Benalaxyl-M	CC1=CC=CC(=C1N(C(C)C(=O)OC)C(=O)CC2=CC=CC=C2)C	36.0	124	77.0	88.0	14
Carbamoyl guanidine (IN-NC148)	COC(=O)C1=C(C=CC=C1)S(=O)(=O)NC(=O)N/C(=N/C(=O)N)/N	12.1	264	78.8	252	16
Terbuthylazine	CCNC1=NC(=NC(=N1)NC(C)(C)C)Cl	58.4	455	112	397	21
Cyproconazole	CC(C1CC1)C(CN2C=NC=N2)(C3=CC=C(C=C3)Cl)O	44.8	347	132	302	11
AMPA	C(N)P(=O)(O)O	41.9	1153	134	1111	11
CSCD465008	C1=C(C(=NN1)C(F)F)C(=O)N	40.4	404	147	364	18

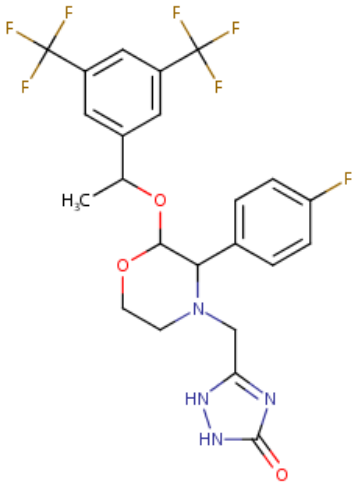
DE-535 pyridinol	C1=C(C(=NC=C1C(F)(F)F)O)Cl	38.0	508	153	470	13
Triazine amine (IN-A4098)	CC1=NC(=NC(=N1)N)OC	7.20	1000	154	993	30
Difenoconazole	CC1COC(CN2C=NC=N2)(C3=CC=C(C=C3Cl)OC4=CC=C(C=C4)Cl)O1	53.0	709	162	656	14
Penconazole	CCCC(CN1C=NC=N1)C2=CC=C(C=C2Cl)Cl	61.0	524	163	463	10
Picloram	C1(=C(C(O)[O-])N=C(C(=C1N)Cl)Cl)Cl	5.00	1451	172	1446	19
Metazachlor acid (BH479-4)	CC1=CC=CC(=C1N(CN2C=CC=N2)C(=O)C(=O)O)C	47.5	589	188	542	16
Methoxypyridine (DMP)	COC1=NC(=C(C(=C1Cl)N)Cl)F	16.7	1000	203	983	12
Chlorantraniliprole	CC1=C(C(=CC(=C1)Cl)C(=O)NC)NC(=O)C2=CC(=NN2C3=NC=CC=C3Cl)Br	20.0	493	214	473	16
Sintofen	COCCOC1=C2C(=CC=C1)N(C3=CC=C(C=C3)Cl)N=C(C2=O)C(=O)O	130	562	221	432	10
Fluopyram	C1=CC(=C(C=C1)C(F)(F)F)C(=O)NCCC2=NC=C(C=C2Cl)C(F)(F)F	162	654	241	492	10
Dimoxystrobin, E isomer more prevalent	CC1=CC=C(C)C(=C1)OCC2=C(C=CC=C2)/C(=N\OC)/C(=O)NC	88.0	1203	301	1115	10
Triticonazole RPA 400727	CC1(C)CC/C(=C/C2=CC=C(C=C2)Cl)/C1(CN3C=NC=N3)O	145	993	334	848	12
Fludioxonil	C1=CC(=C2C(=C1)OC(F)(F)O2)C3=CNC=C3C#N	79.0	832	342	753	15
Epoxiconazole	C1=CC(=C(C=C1)Cl)C2C(CN3C=NC=N3)(C4=CC=C(C=C4)F)O2	161	2113	453	1952	17
Flutriafol	C1=CC(=C(C=C1)F)C(CN2C=NC=N2)(C3=CC=C(C=C3)F)O	672	3492	1696	2820	10

Table S3. Examples of pharmaceuticals that fall outside the *Eawag-BBD* domain but inside the *Eawag-Soil* domain, and the three most similar compounds (nearest neighbors) in the *Eawag-Soil* data set.

Pharmaceutical		
		
1st nearest neighbor in *Eawag-Soil*	2nd nearest neighbor in *Eawag-Soil*	3rd nearest neighbor in *Eawag-Soil*
		
Pharmaceutical		



1st nearest neighbor in *Eawag-Soil*	2nd nearest neighbor in *Eawag-Soil*	3rd nearest neighbor in *Eawag-Soil*
		
Pharmaceutical		



1st nearest neighbor in *Eawag-Soil*	2nd nearest neighbor in *Eawag-Soil*	3rd nearest neighbor in *Eawag-Soil*