Supporting Information

for

Insecticide Risk in United States Surface Waters: Drivers and Spatiotemporal Modeling

Jakob Wolfram[†], Sebastian Stehle[†,‡], Sascha Bub[†], Lara L. Petschick[†], Ralf Schulz[*,†]

[†]iES Landau, Institute for Environmental Sciences, University of Koblenz-Landau, Fortstrasse 7, D-76829 Landau, Germany

[‡]Eusserthal Ecosystem Research Station, University of Koblenz-Landau, Birkenthalstrasse 13, D-76857 Eusserthal, Germany

[*] Phone: +49634128031327; e-mail: schulz@uni-landau.de

Pages S1 – S11

Tables S1 – S7

Figures S1 – S3

Methods:

List of degradation products

In the following, a list of all degradation products that were deemed to be potentially relevant to aquatic environments, is provided and more detailed descriptions for their inclusion in this meta-analysis can be found in Wolfram et al.[1]:

Endosulfan sulfate, azinphos-methyl oxygen analog, 3,5,6-trichloro-2-pyridinol, chlorpyrifos oxygen analog, chlorpyrifos oxon, diazoxon, diazinon oxon, malaoxon, isomalathion, malathion dicarboxylic acid, malathion monocarboxylic acid, paraoxon, methyl paraoxon, 3-hydroxycarbofuran, 3-ketocarbofuran, carbofuran phenol, alpha-R-deltamethrin, 6-chloronicotinic acid, imidacloprid guanidine, imidacloprid olefin, imidacloprid olefinic-guanidine, imidacloprid urea, o-p-dichlorobenzophenone, p-p-dichlorobenzophenone, 1,1-bis(p-chlorophenyl-)2,2-dichloroethanol, dichlorobenzhydrol, hydroxyl-dichlorobenzophenone, aldicarb sulfoxide, aldoxycarb, aldicarb sulfone, N-(2-chloro-5-thizolylmethyl)-N'-nitroguanidine, 1-methyl-2-nitro-3-(tetrahydro-3-furylmethyl)-guanidine, 1-methylguanidinium chloride, 3-(methylamino)-9-oxa-2-aza-4-azoniabicyclone-3-enehydrogen succinate, 1-methyl-2-nitroguanidine, 1-methyl-3-(tetrahydro-3-furylmethyl)guanidine, 1-methyl-3-(tetrahydro-3-furylmethyl)-guanidinium dihydrogen phosphate, fipronil sulfone, fipronil sulfide, and desulfinyl fipronil.

Multiple linear regression validation

MLRs were validated following suggestions of Unwin[2] and references therein. Briefly, the normality distribution of the residuals was checked visually using histograms and QQ-Plots, while the independence of errors was tested using the Durbin-Watson-Test with $1>d>3$ representing thresholds that would suggest autocorrelation.[3] Homoscedasticity was assessed visually by plotting standardized residuals versus their fitted values. Additionally, the

significance of all regression coefficients was checked using White's heteroscedasticity corrected covariance matrix[4] and by calculating regression coefficient 95% confidence intervals. Both tests did not reveal any departure from the initial MLR results. Outliers were assessed with studentized residuals and standardized residuals, while potential leverage points were assessed using hat-values, DFFit, Cook's distance and covariance ratios. Multicolinearity was assessed using variance inflation factors with a threshold level of 3. For MLRs using factorial attributes with more than one degree of freedom (df), generalized variance inflation factors were calculated (GVIF, see Fox and Monette[5]). GVIFs were then transformed back via squaring, allowing comparison with traditionally used threshold levels (i.e. VIF < 3). Only substances with more than 20 MICs were used for all MLRs, and no violations of concern became apparent.

Internal validation via simulation

Following aforementioned standard validation exercises, every model was further tested for robustness and validity applying internal validation techniques,[6,7] using a simulation approach (i.e. bootstrapping , n = 10,000) of randomly drawn data (with replacements).

Surface water risk (individual substances)

Detailed summary statistics about bootstrapping simulation runs (n = 10,000) of the linear model for individually measured MICs in freshwaters (n = 1,833, R^2 adj. = 62.17, $p < 0.001$) are provided in Table S3. Bootstrapping simulation runs (n = 10,000) and their resulting adjusted R^2 are summarized in Figure S3.

Sediment risk (individual substances)

Detailed summary statistics about bootstrapping simulation runs (n = 10,000) of the linear model for individually measured MICs in sediments (n = 478, R^2 adj. = 48.10, $p < 0.001$) are

provided in Table S6. Bootstrapping simulation runs (n = 10,000) and their resulting adjusted R^2 are summarized in Figure S2.

Surface water risk (maxTU)

Detailed summary statistics about bootstrapping simulation runs (n = 10,000) of the maximum toxicant pressure (i.e. maxTU) model for surface waters (n = 496, R^2 adj. = 75.17, $p < 0.001$) are provided in Table S7. Bootstrapping simulation runs (n = 10,000) and their resulting adjusted R^2 are summarized in Figure S3.

References

1. Wolfram, J.; Stehle, S.; Bub, S.; Petschick, L. L.; Schulz, R., Meta-Analysis of Insecticides in United States Surface Waters: Status and Future Implications. *Environ. Sci. Technol.* **2018,** *52*, (24), 14452-14460.
2. Unwin, A., Discovering Statistics Using R by Andy Field, Jeremy Miles, Zoë Field. *International Statistical Review* **2013,** *81*, (1), 169-170.
3. Durbin, J.; Watson, G. S., Testing for serial correlation in least squares regression. II. *Biometrika* **1951,** *38*, (1/2), 159-177.
4. White, H., A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. *Econometrica* **1980,** *48*, (4), 817-838.
5. Fox, J.; Monette, G., Generalized collinearity diagnostics. *Journal of the American Statistical Association* **1992,** *87*, (417), 178-183.
6. Harrell Jr, F. E., *Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis*. Springer: 2015.
7. Steyerberg, E. W.; Harrell Jr, F. E.; Borsboom, G. J.; Eijkemans, M.; Vergouwe, Y.; Habbema, J. D. F., Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. *J. Clin. Epidemiol.* **2001,** *54*, (8), 774-781.
8. Lewis, K. A.; Tzilivakis, J.; Warner, D. J.; Green, A., An international database for pesticide risk assessments and management. *Human and Ecological Risk Assessment: An International Journal* **2016,** *22*, (4), 1050-1064.

List of Tables

Table S1: Regulatory threshold levels (RTL) for freshwater (RTL_{fw}), sediments (RTL_{sed}) and estuaries (RTL_{est}) for all 32 parent compounds. Only insecticide degradates with reported concentrations are listed and provided with a RTL. A full list of degradates of potential environmental concern can be found in the first paragraph of the SI. Please note, this table can also be found in Wolfram et al.[1]

Insecticide class	Insecticide	RTL_{fw} (µg/L)	RTL_{sed} (µg/kg)	RTL_{est} (µg/L)
Organochlorine	Dicofol	26.5	3680	7.55
	Endosulfan	0.05	3.84	0.02
Organophosphate	Azinphos-methyl	0.08	0.89	0.105
	Chlorpyrifos	0.03	16	0.0175
	Diazinon	0.105	76	2.1
	Malathion	0.295	2.56	1.1
	Parathion-ethyl	0.02	0.325	0.0535
	Parathion-methyl	0.485	20.8	0.175
Carbamate	Aldicarb	10		6
	Carbaryl	0.85	4.4	2.85
	Carbofuran	1.115	1.72	2.3
Pyrethroid	Bifenthrin	0.0002465	0.25	0.001985
	Cyfluthrin	0.0125	0.53	0.0012
	Cypermethrin	0.0018	0.308	0.002375
	ζ-cypermethrin	0.0018	0.308	0.002375
	Deltamethrin	0.0001	0.48	0.00185
	Esfenvalerate	0.000424	7.4	0.00233
	Fenpropathrin	0.001525	12.4	0.0105
	Fenvalerate	0.016	2.2	0.004
	λ-cyhalothrin	0.00015	0.31	0.00245
	Permethrin	0.0033	7.4	0.009
	Resmethrin	0.14		0.115
	τ-fluvalinate	0.155		0.003
	Tefluthrin	0.03	11.6	0.0265
	Tetramethrin	1.85		1.85
	Tralomethrin	0.0195		0.4225
Neonicotinoid	Acetamiprid	10.5		33
	Clothianidin	11		26.5
	Dinotefuran	49550		395
	Imidacloprid	0.385		16.5
	Thiamethoxam	17.5	10	3450
Phenylpyrazole	Fipronil	0.11	16	0.07
Degradate	Endosulfan sulfate	0.05	73	0.02
	Aldicarb sulfoxide	21.5		
	Aldicarb sulfone	184.5		
	Fipronil sulfone	0.36	9.1	0.28

Fipronil sulfide	1.065	29	0.0385
Desulfinyl fipronil	100	200	0.75

Table S2: List of covariates and attributes with their respective data type and availability in percent that were extracted from publications and assigned to individual MICs.

Covariate	Data type	Description	Available %
Substance name	factor	Name of detected insecticide	100
Chemical class	factor	Insecticide class (e.g. Organophosphate)	100
TU identifier	integer	Shared identifier per sample	100
Reference	factor	Name and date of the respective publication	100
Compartment	integer	Integer detailing if MIC was detected in the water phase (1) or sediment phase (2)	100
Sampling interval	numeric	Time in days between individual samples per location	77.4
Catchment size	numeric	Size in km² of contributing catchment area per sampling location	81.7
Sampling date	numeric	Date of sampling event	98.1
Location name	factor	Name of sampling location	100
Agricultural produce	factor	List of crops grown in respective catchment	71.2
Sampling location (state)	factor	Federal state in which the sampling location is situated	97.8
Smapling location (county)	factor	County in which the sampling location is situated	78.7
Water body type	factor	Water body type (e.g. creek, estuary, river) as detailed in the publication	95.6
Freshwater	integer	Integer detailing if the respective water body contains freshwater (1) or estuarine water (2)	100
Flow type	integer	Interger detailing if respective water body is flowing (1) or standing (2)	97
Number of detects	integer	Number of detections per substance as detailed in the publication	97.7
Numbe rof non-detects	integer	Number of non-detections per substance as detailed in the publication	83.1
Filtration	factor	Description of filtration procedures used for pyrethroid samples	95.7
LOD	numeric	Analytical limit of detection for respective substance	39.9
LOQ	numeric	Analytical limit of quantification for respective substance	51.1
Additional pesticides	factor	Factor indicating if additional pesticides (excluding insecticides) were analyzed	97.7
Additional insecticides	factor	Factor indicating if additional insecticides were analyzed	100
Total pesticides	integer	Number of total detected pesticides (excluding insecticides) per sample	100
Total insecticides	integer	Number of total detected insecticides per sample	65.2
Total Pesticides/Insecticides	factor	Factor detailing how many pesticides/insecticides were detected	99
Detected pesticides	factor	List of susbtance names detected per sample	61.7
Agricultural origin	factor	Factor describing certainty that MICs originated from agricultural NPS as detailed in the publication	100

Agricultural proximity	factor	Factor describing if any information was provided regarding the proximity of agricultural fields to the sampling location	60.9
Highest spatial accuracy	integer	Integer detailing if MIC was attributable to a location (1), county (2), or the federal state (3)	100
Latitude	numeric	Latitude of sampling location (GPS)	75.3
Longitude	numeric	Longitude of sampling location (GPS)	75.3
RTL	numeric	respective regulatory threshold level	100
RTL class	integer	Integer detailing if MIC was attributable to a freshwaters (1), sediments (2), or estuarine waters (3)	100

Table S3: Detailed summary statistics of model validation simulation runs (bootstrapping, n = 10,000) and resulting bootstrapped 95% confidence intervals for the freshwater linear regression of individual MIC to RTL ratios. Standardized betas with their respective rank (SBR) and variance inflation factors (VIF) of the main model are provided for reference.

Predictor variable	Lower 95%-confidence limit	2nd quartile (median)	Upper 95%-confidence limit	Sign.[1]	Std. beta	SBR[2]	VIF
Intercept	-2.302	-1.928	-1.542	*			
Catchment size [km²][3]	-0.183	-0.151	-0.120	*	-0.146	4	1.897
Sampling interval [d][3]	-0.227	-0.199	-0.170	*	-0.189	3	1.311
Time [y]	-0.064	-0.057	-0.050	*	-0.287	2	2.283
toxicity-normalized use [kg × RTL^{-1}][3]	0.350	0.385	0.422	*	0.381	1	1.987
Burst factor	0.011	0.015	0.019	*	0.131	5	2.310
Irrigated agricultural land-use [%]	0.289	0.427	0.560	*	0.101	6	1.448
Organophosphates	-0.034	0.242	0.518				1.653
Pyrethroids	0.944	1.231	1.516	*			1.653
Neonicotinoids	0.130	0.436	0.726	*			1.653
Phenylpyrazole	0.807	1.235	1.678	*			1.653

[1] Determined based on confidence level overlap

[2] Ranked standardized betas of linear model

[3] Predictor variable was transformed by the decadal logarithm

Table S4: Comparison of insecticides' stability in days (i.e. DT_{50}) in different compartments and the resulting delta in days. Data was obtained from Lewis et al.[8].

Substance	DT_{50} in days (sediment)	DT_{50} in days (water)	Delta
Malathion	0.4	0.4	0
Diazinon	10.4	4.3	6.1
Endosulfan	*N/A*	*N/A*	*N/A*
Chlorpyrifos	36.5	5	31.5
Bifenthrin	161	8	153
ʎ-cyhalothrin	15.1	0.24	14.86

Esfenvalerate	56	30	26
Fenpropathrin	28	1	27
Carbaryl	5.8	3.1	2.7
Carbofuran	9.7	6.1	3.6
Permethrin	40	23	17
Deltamethrin	65	17	48
Cypermethrin	17	3	14
Fenvalerate	*N/A*	*N/A*	*N/A*
Parathion-methyl	4.3	3.5	0.8
Cyfluthrin	8	1	7
Cypermethrin-zeta	2	0.1	1.9
Fipronil	68	54	14
Desulfinyl-Fipronil	*N/A*	*N/A*	*N/A*
Fipronil sulfide	*N/A*	*N/A*	*N/A*
Fipronil sulfone	*N/A*	*N/A*	*N/A*
Dicofol	29	*NA*	*N/A*

Table S5: Toxic modes of action for respective insecticide classes used in the present study.

Insecticide class	Toxic Mode of Action (TMoA)
Organochlorine	GABA-gated chloride channel antagonist
Organophosphate	Acetylcholinesterase inhibitor
Carbamate	Acetylcholinesterase inhibitor
Pyrethroid	Sodium channel modulators
Neonicotinoid	Nicotinic acetylcholine receptor agonists
Phenylpyrazole	$GABA_A$-gated chloride channel antagonist

Table S6: Detailed summary statistics of model validation simulation runs (bootstrapping, n = 10,000) and resulting bootstrapped 95% confidence intervals for the sediment linear regression of individual MIC to RTL ratios. Standardized betas with their respective rank (SBR) and variance inflation factors (VIF) of the main model are provided for reference.

Predictor variable	Lower 95%-confidence limit	2nd quartile (median)	Upper 95%-confidence limit	Sign.[1]	Std. beta	SBR[2]	VIF
Intercept	-3.239	-2.414	-1.602	*			
Catchment size [km²]²	-0.142	-0.101	-0.061	*	-0.097	6	1.150
Sampling interval [d]²	-0.085	-0.053	-0.021	*	-0.075	7	1.257
Time [y]	-0.085	-0.067	-0.049	*	-0.215	2	1.274
Toxicity-normalized use [kg × RTL-1]²	0.087	0.188	0.281	*	0.170	3	1.808
Lentic bodies [factor]	-0.528	-0.367	-0.210	*	-0.098	5	1.336
Irrigated agricultural land-use [%]	0.312	0.481	0.656	*	0.126	4	1.236
KOC [L/kg]²	0.617	0.769	0.965	*	0.506	1	2.879
Organophosphates [factor]	-1.326	-0.783	-0.301	*			1.368

Pyrethroids [factor]	-0.946	-0.421	0.063		1.368
Phenylpyrazole [factor]	-2.117	-0.798	0.379		1.368

[1] Determined based on confidence level overlap

[2] Predictor variable was transformed by the decadal logarithm

[3] Ranked standardized betas of linear model

Table S7: Detailed summary statistics of model validation simulation runs (bootstrapping, n = 10,000) and resulting bootstrapped 95% confidence intervals for the mixture toxicity (maxTU, freshwaters) linear regression of individual MIC to RTL ratios. Standardized betas with their respective rank (SBR) and variance inflation factors (VIF) of the main model are provided for reference.

Predictor variable	Lower 95%-confidence limit	2nd quartile (median)	Upper 95%-confidence limit	Sign.[1]	Std. beta	SBR[2]	VIF
Intercept	-2.219	-1.736	-1.243	*			
Catchment size [km²][3]	-0.393	-0.314	-0.238	*	-0.246	3	2.071
Sampling interval [d][3]	-0.194	-0.130	-0.063	*	-0.106	7	1.470
Burst factor	0.010	0.018	0.025	*	0.138	6	2.180
Time [y]	-0.074	-0.063	-0.053	*	-0.269	2	1.180
Toxicity-normalized use [kg × RTL^{-1}][3]	0.394	0.448	0.503	*	0.411	1	1.490
Number of detected substances	0.050	0.073	0.097	*	0.151	5	1.665
Irrigated agricultural land-use [%]	0.726	0.989	1.254	*	0.209	4	1.453

[1] Determined based on confidence level overlap

[2] Ranked standardized betas of linear model

[3] Predictor variable was transformed by the decadal logarithm

List of figures

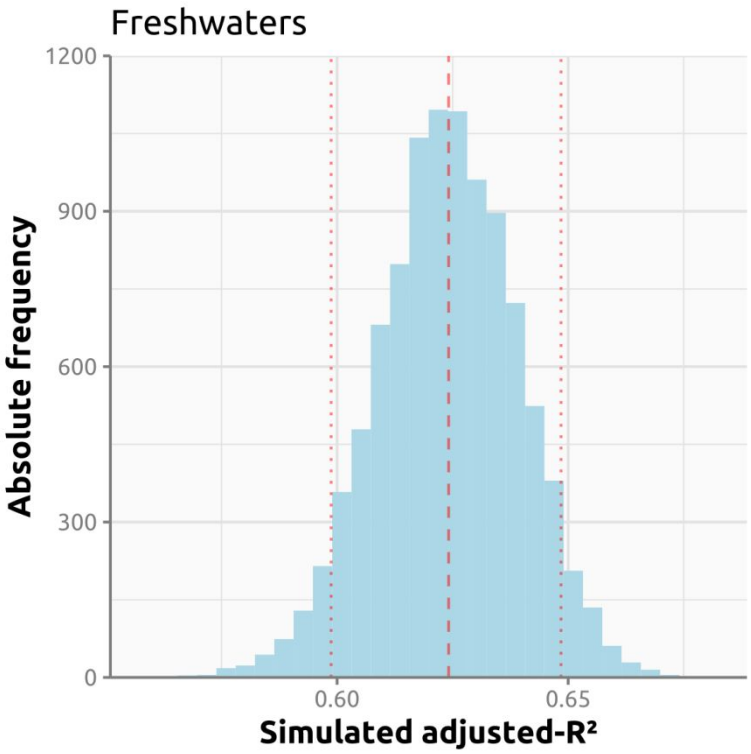

Figure S1: Frequency plot detailing the distribution of bootstrap simulated (n = 10,000) adj. R^2 of the multiple linear regression for insecticide risks in freshwaters. Bootstrapped upper and lower 95% confidence intervals are depicted by dotted (orange) lines and the median is depicted by a red dashed line.

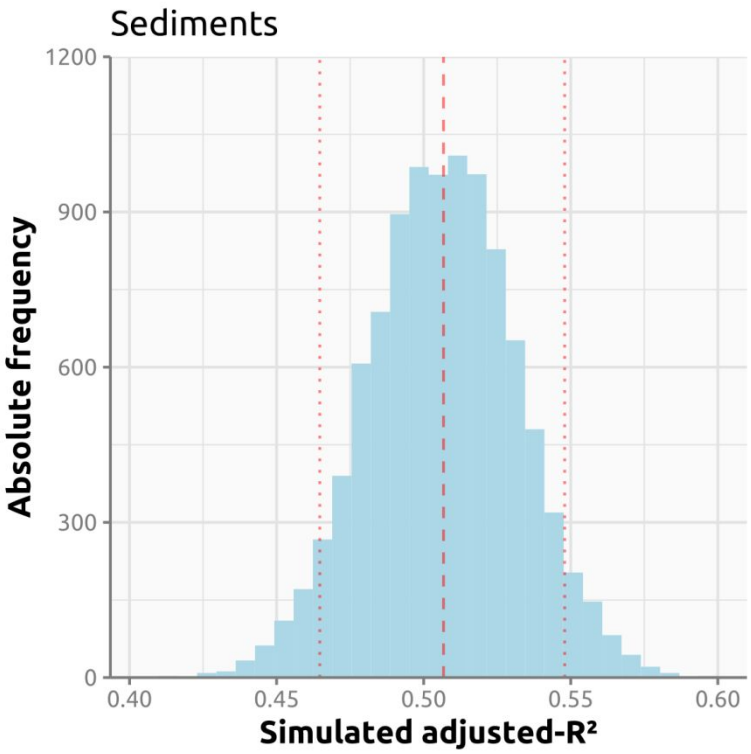

Figure S2: Frequency plot detailing the distribution of bootstrap simulated (n = 10,000) adj. R^2 of the multiple linear regression for insecticide risks in sediments. Bootstrapped upper and lower 95% confidence intervals are depicted by dotted (orange) lines and the median is depicted by a red dashed line.

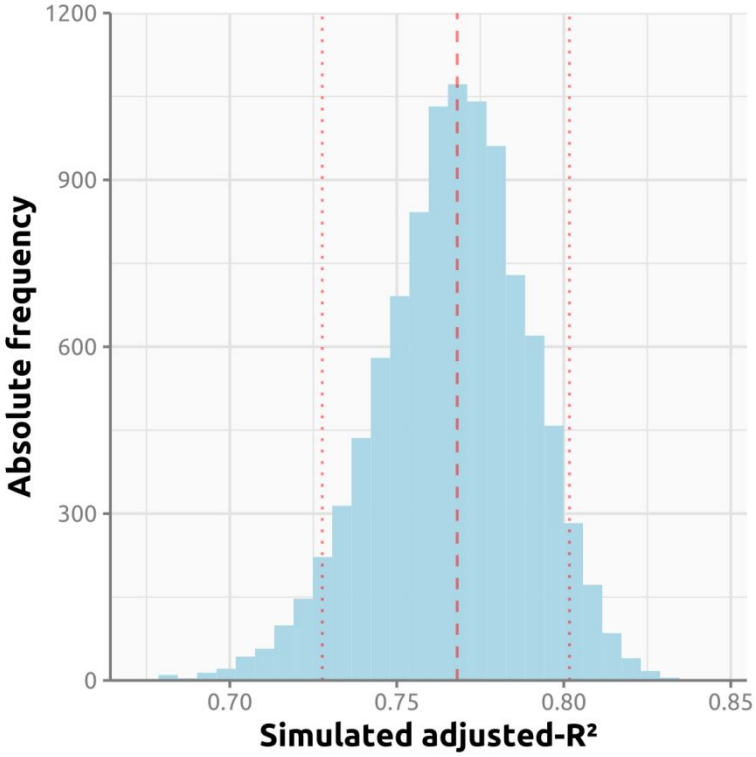

Figure S3: Frequency plot detailing the distribution of bootstrap simulated (n = 10,000) adj. R^2 of the maxTU multiple linear regression. Bootstrapped upper and lower 95% confidence intervals are depicted by dotted (orange) lines and the median is depicted by a red dashed line.