A regulatory perspective on the systematic use of high-resolution mass spectrometry non-target screening data in environmental monitoring and chemicals management – use cases from the German NTSPortal

Additional file 1

Anna Lena Kronsbein[1], Ronya Mona Wallis[1], Eric Winter[1], Nicole Bandow[1], Kevin S. Jewell[2], Georg Dierkes[2], Arne Wick[2], and Jan Koschorreck[1]*

*corresponding author

[1] German Environment Agency (Umweltbundesamt), Wörlitzer Platz 1, 06813 Dessau-Roßlau, Germany

[2] Federal Institute of Hydrology (Bundesanstalt für Gewässerkunde), Am Mainzer Tor 1, 56068 Koblenz, Germany

S-1 Methods

S-1.1. EU market list and collective spectral library

The categorization of substances in the CSL was performed using a multi-step approach in Python. First the CSL was merged to the market list (see Table S 1-1 for data sources). The market list includes all substances present in the original data, with information on the approval or registration status, such as “approved”, “pending for approval”, “no longer approved”, “not approved” and others.

CPDat was additionally used for the categorization. Categories in CPDat were aggregated using the following scheme: Home, Furniture, Raw Material, Vehicle and Construction as Industrial Substances, cosmetics and personal care products as Cosmetics, herbicides and insecticides as Biocidal Products/PPP. The legal status of substances classified under CPDat cannot be conclusively clarified. Further steps are described in Methods (LC-HRMS measurements, collective spectral library and quality control).

Table S 1-1: EU legislations, their data source and date of download, used to assign the substances on the market list to the legislation under which they were placed on the market.

Legislation	Data Source	Time of Download
Biocidal Products	https://echa.europa.eu/de/information-on-chemicals/biocidal-active-substances	23.05.2025

PPP	https://ec.europa.eu/food/plant/pesticides/eu-pesticides-database/start/screen/active-substances	23.05.2025
REACH	https://echa.europa.eu/de/information-on-chemicals/registered-substances data remains frozen as of 19 May 2023	23.05.2025
Cosmetics	https://echa.europa.eu/de/cosmetics-colorant https://echa.europa.eu/de/cosmetics-preservatives https://echa.europa.eu/de/cosmetics-restricted-substances https://echa.europa.eu/de/cosmetics-uv-filters	23.05.2025
Food Contact Materials	https://echa.europa.eu/de/plastic-material-food-contact	23.05.2025
Feed Additives	https://ec.europa.eu/food/food-feed-portal/screen/home	23.05.2025
Food Additives	https://ec.europa.eu/food/food-feed-portal/screen/food-additives/search	23.05.2025
Food Flavouring	https://ec.europa.eu/food/food-feed-portal/screen/food-flavourings/search	23.05.2025
Medicinal Products EMA	https://www.ema.europa.eu/en/medicines/download-medicine-data	23.05.2025
Medicinal Products HMA	https://mri.cts-mrp.eu/portal/advanced-search	23.05.2025
Medicinal Products BFARM	https://portal.dimdi.de/amguifree/am/search.xhtml https://www.bfarm.de/DE/Arzneimittel/Arzneimittelinformationen/Arzneimittel-recherchieren/Stoffbezeichnungen/_node.html	03.12.2024

S-1.2. Samples and sample preparation

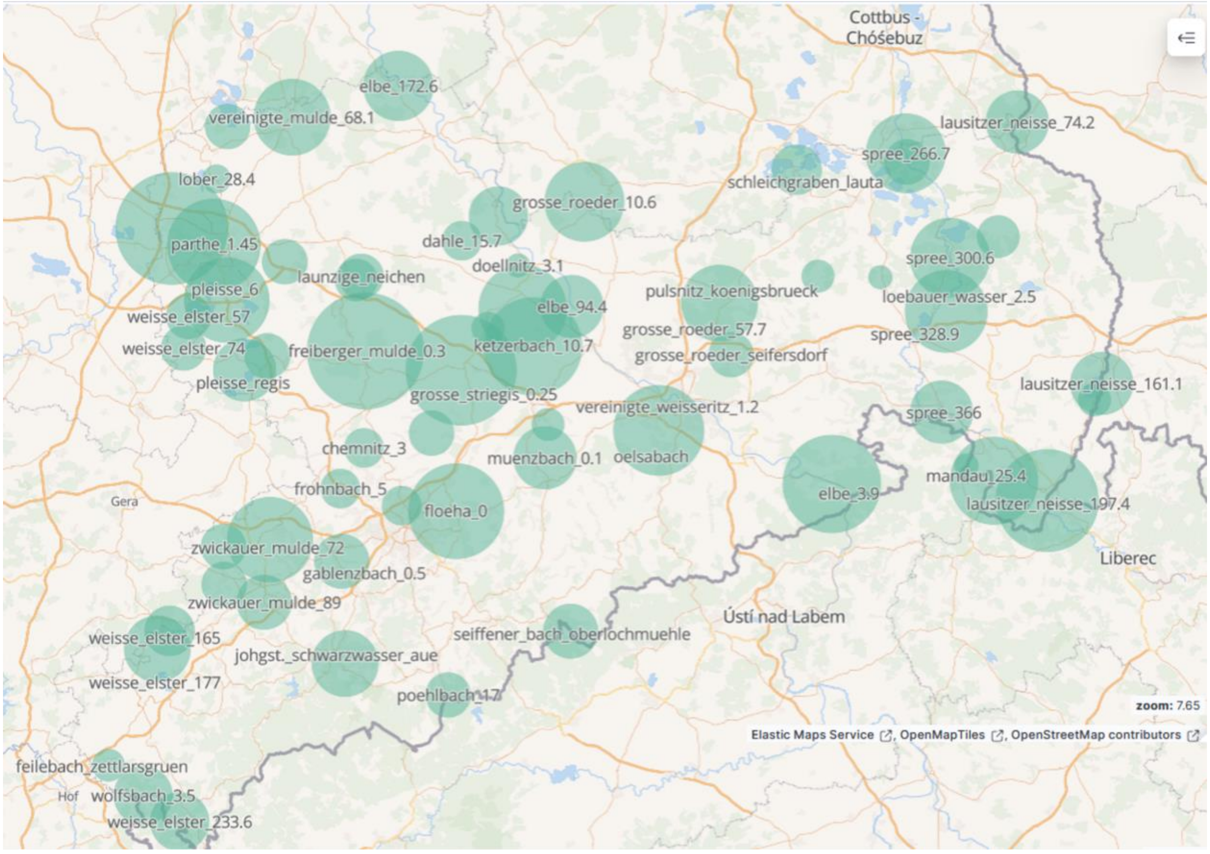

Figure S 1-1: Map of surface water sampling sites in Saxony, Germany. Circle size presents the number of measurements. Screenshot taken from the online user interface of the NTSPortal.

S-1.3. LC-HRMS measurements, collective spectral library and quality assurance

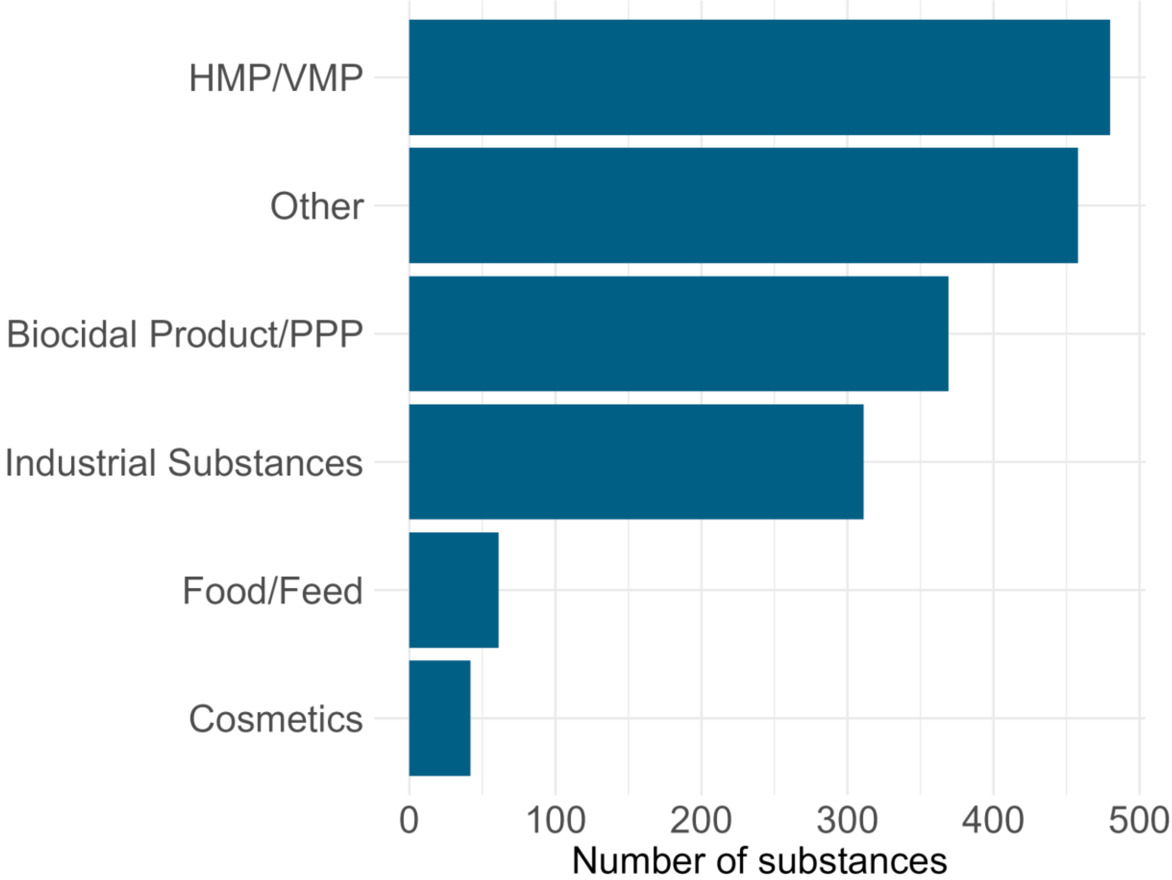

Figure S 1-2: Distribution of 1721 collective spectral library substances among the individual assigned groups.

S-1.4. Data analysis

Annual geometric mean normalised peak areas were used as the response variable in generalized linear models (GLMs) and generalized additive models (GAMs). The variable river was included as a categorical factor to account for site-specific effects. GLMs assumed a linear relationship between peak areas and year, whereas GAMs allowed for non-linear temporal trends using penalized regression splines. GAMs were fitted under three distributional assumptions - Gaussian, Gamma (log link), and Tweedie (log link) - using restricted maximum likelihood (REML) estimation. Model performance was evaluated using the Akaike Information Criterion (AIC) and residual diagnostics based on Monte Carlo simulations (DHARMa). The model with the lowest AIC and no over or underdispersion was selected as the best-fitting representation of the temporal trend. For visualization and summary statistics, predictions were generated from the best model on a regular grid covering the full range of observation years. Predictions were computed separately for each river and then combined into a weighted marginal smooth, with each river's contribution weighted by its number of observations. This approach provided both river-specific and overall (across-river) fitted trend lines with corresponding confidence intervals.

S-2 Results and discussion

S-2.1. Spatial distribution of chemical mixtures

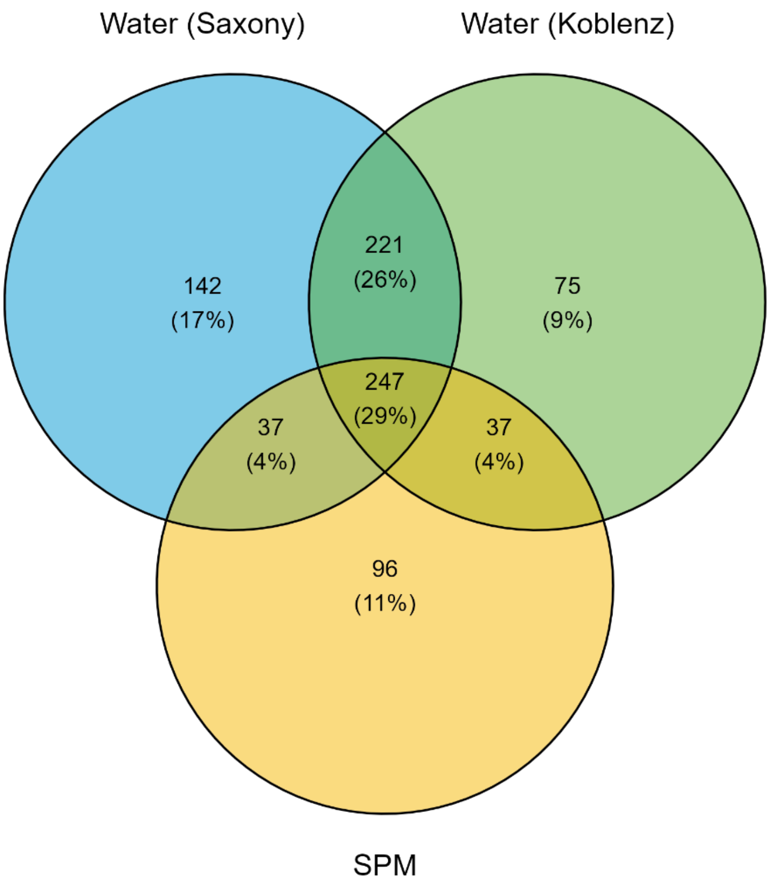

Figure S 2-1: Venn diagram showing the intersections of substances detected in different NTSPortal datasets. Water samples from Koblenz: green, water samples from Saxony: blue, annual suspended particulate matter samples: yellow.

S-2.2. Temporal changes in mixture composition

Table S 2-1: Number of detected substances at earliest (D1, D3: 2009; S1, S2: 2006; all other stations: 2005) and latest (2022) annual SPM sampling timepoint at all sampling stations.

SPM station	Earliest timepoint	Latest timepoint	Difference	Direction
D1	125	124	-1	Decrease
D3	125	115	-10	Decrease
E1	142	155	13	Increase
E2	165	158	-7	Decrease
E4	117	116	-1	Decrease
E5	156	155	-1	Decrease
Mu	157	151	-6	Decrease
R1	128	114	-14	Decrease
R3	181	179	-2	Decrease
R4	205	193	-12	Decrease

S1	172	170	-2	Decrease
S2	182	191	9	Increase
Sa	182	190	8	Increase

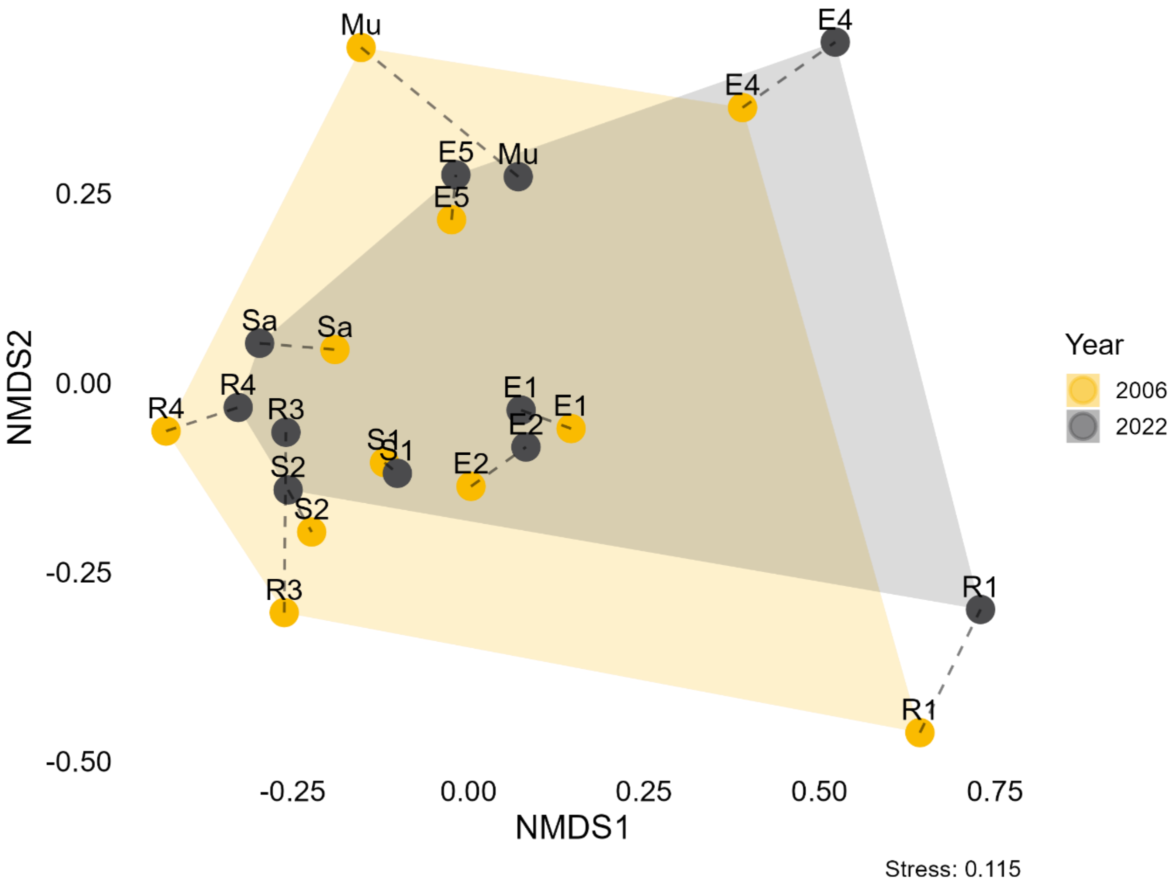

Figure S 2-2: Non-metric multidimensional scaling (NMDS) plot of Jaccard distances of substances detected in annual SPM samples from eleven stations (E1, E2, E4, E5, Mu, R1, R3, R4, S1, S2, Sa) in 2006 (orange) and 2022 (grey). Dashed lines connect the same stations in different years. Shaded areas indicate the outline of station points for each year, showing overall group distribution. Stress value = 0.115.

S-2.3. Environmental legislation - EU WFD and Watch List

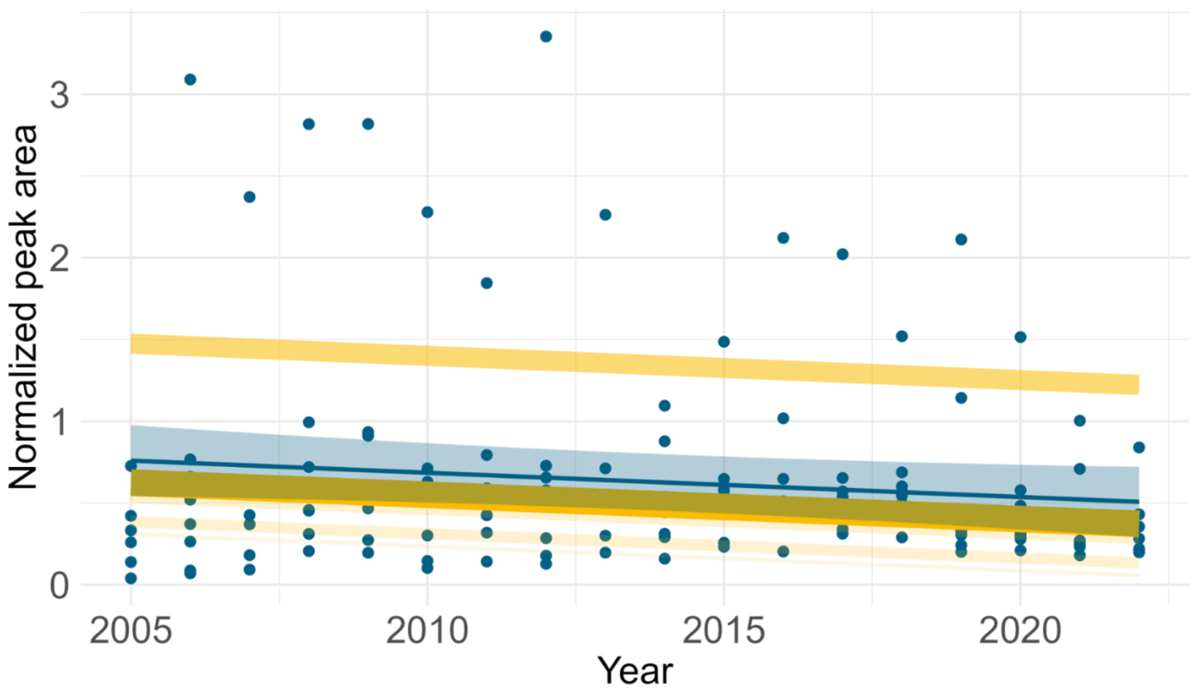

Figure S 2-3: *Normalized peak areas of 6PPD-quinone in SPM samples from the Elbe (E2), Mulde (Mu), Rhine (R1, R3, R4), Saale (Sa), and Saar (S1, S2) rivers. Blue points represent individual annual samples. Blue line shows the overall GLM fit with a weighted smooth and 95% confidence interval (p(year) = 0.05, slope = – 0.01). Orange lines indicate river-specific contributions, with darker and thicker lines reflecting stronger influence.*

S-2.4. Chemical legislation – Cosmetic Products Regulation

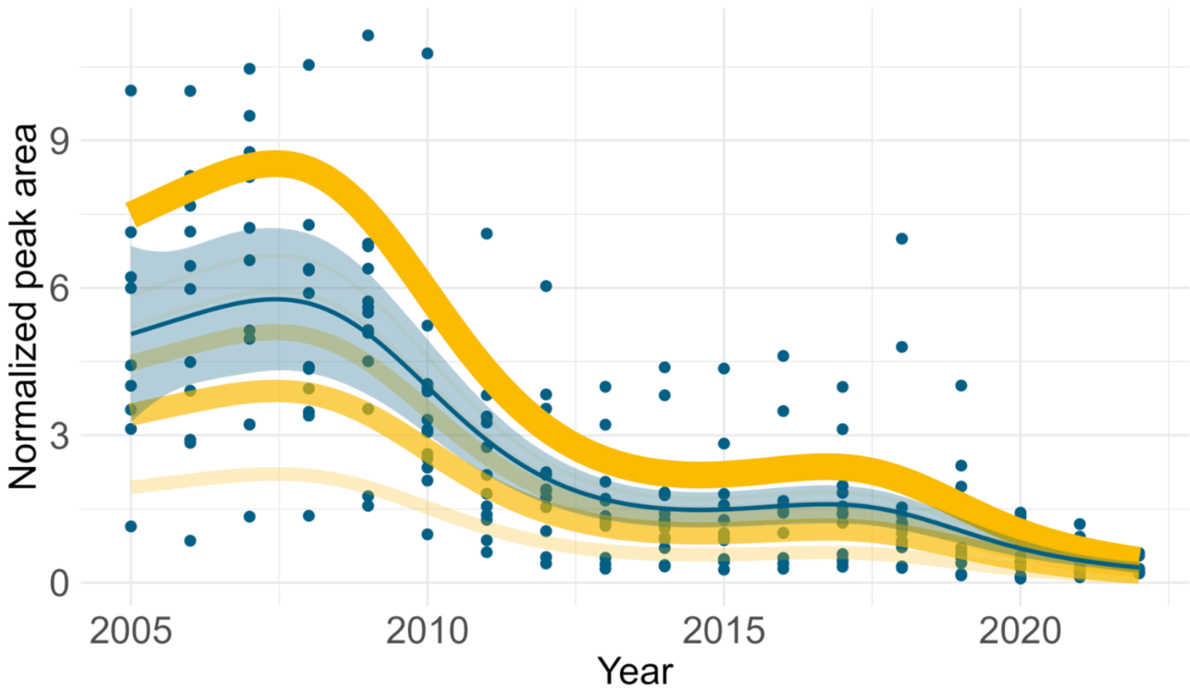

Figure S 2-4: Normalized peak areas of Climbazole in SPM samples from the Elbe (E1, E2, E4, E5), Mulde (Mu), Rhine (R1, R3, R4), Saale (Sa), Saar (S1, S2), and Danube (D1, D3) rivers. Blue points represent individual annual samples. Blue line shows the overall GAM Gamma fit with a weighted smooth and 95% confidence interval (p(year) < 0.0001, slope = –0.28). Orange lines indicate river-specific contributions, with darker and thicker lines reflecting stronger influence.

S-2.5. Chemical legislation – REACH Regulation and national Centre for Micropollutants

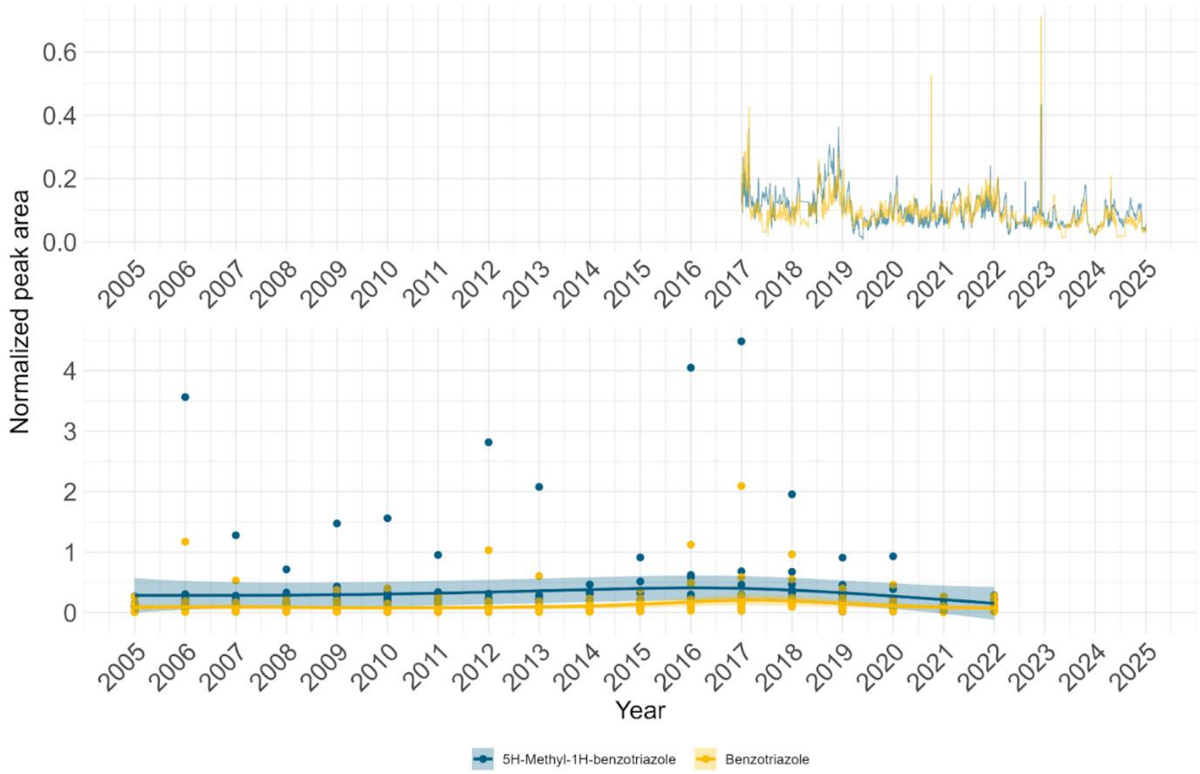

Figure S 2-5: Normalized peak areas of 1H-benzotriazole (orange) and 4- and/or 5-Methyl-1H-benzotriazole (blue) in (upper panel) surface water at Koblenz, expressed as a 7-day rolling mean, and (lower panel) SPM samples from the Elbe (E1, E2, E4, E5), Mulde (Mu), Rhine (R3, R4), Saale (Sa), and Saar (S1, S2) rivers. Blue points indicate individual annual measurements, while the lines represent the GAM fit with a weighted smooth and associated 95% confidence interval. 1H-benzotriazole: GAM Gamma fit, p(year) = 0.001, slope = 0.00; 4 and/or 5-Methyl-1H-benzotriazole: GAM Gaussian fit, p(year) > 0.05, slope = –0.01.*

**In this study, the two regioisomers could not be analytically distinguished. Therefore, any detection of 4-Methylbenzotriazole was reported as representing both 4- and 5-Methylbenzotriazole and vice versa.*

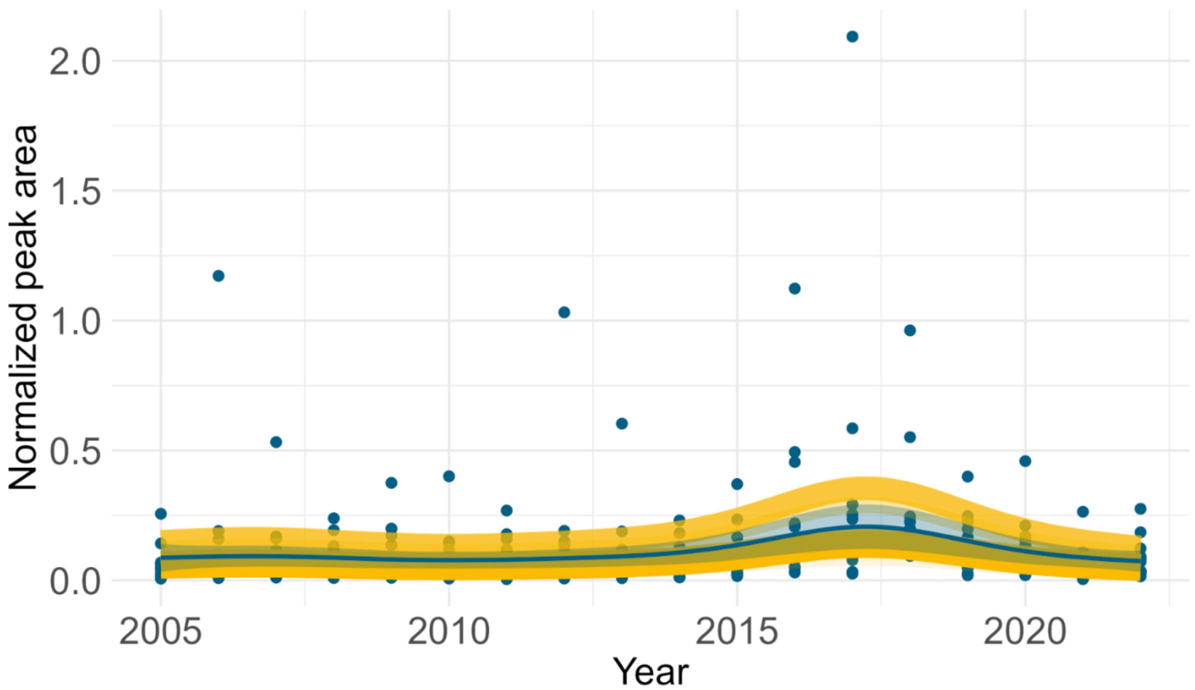

Figure S 2-6: Normalized peak areas of 1H-Benzotriazole in SPM samples from the Elbe (E1, E2, E5), Mulde (Mu), Rhine (R1, R3, R4), Saale (Sa), Saar (S1, S2), and Danube (D1, D3) rivers. Blue points represent individual annual samples. Blue line shows the overall GAM Gamma fit with a weighted smooth and 95% confidence interval (p(year) = 0.001, slope = 0.00). Orange lines indicate river-specific contributions, with darker and thicker lines reflecting stronger influence.

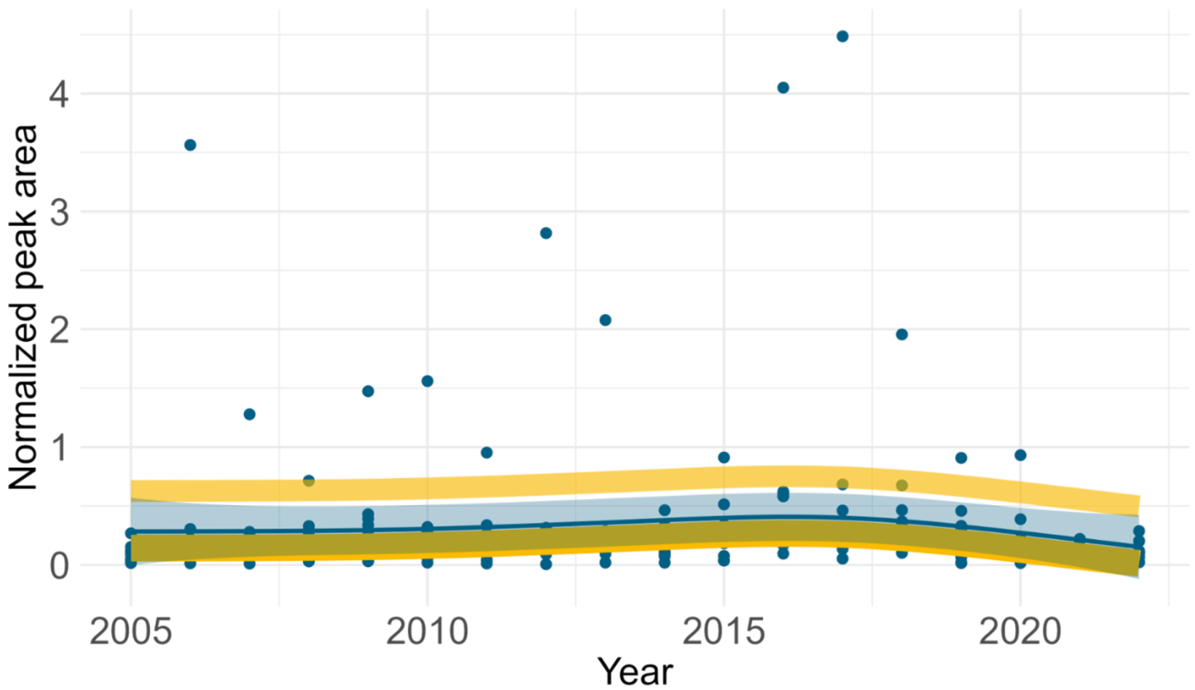

Figure S 2-7: Normalized peak areas of 4- and/or 5-Methyl-1H-benzotriazole in SPM samples from the Elbe (E1, E2, E4, E5), Mulde (Mu), Rhine (R3, R4), Saale (Sa), and Saar (S1, S2) rivers. Blue points represent individual annual samples. Blue line shows the overall GAM Gaussian fit with a weighted smooth and 95%*

confidence interval (p(year) > 0.05, slope = –0.01). Orange lines indicate river-specific contributions, with darker and thicker lines reflecting stronger influence.

**In this study, the two regioisomers could not be analytically distinguished. Therefore, any detection of 4-Methylbenzotriazole was reported as representing both 4- and 5-Methylbenzotriazole and vice versa.*

S-2.6. Chemical legislation – Plant Protection and Biocidal Products Regulation

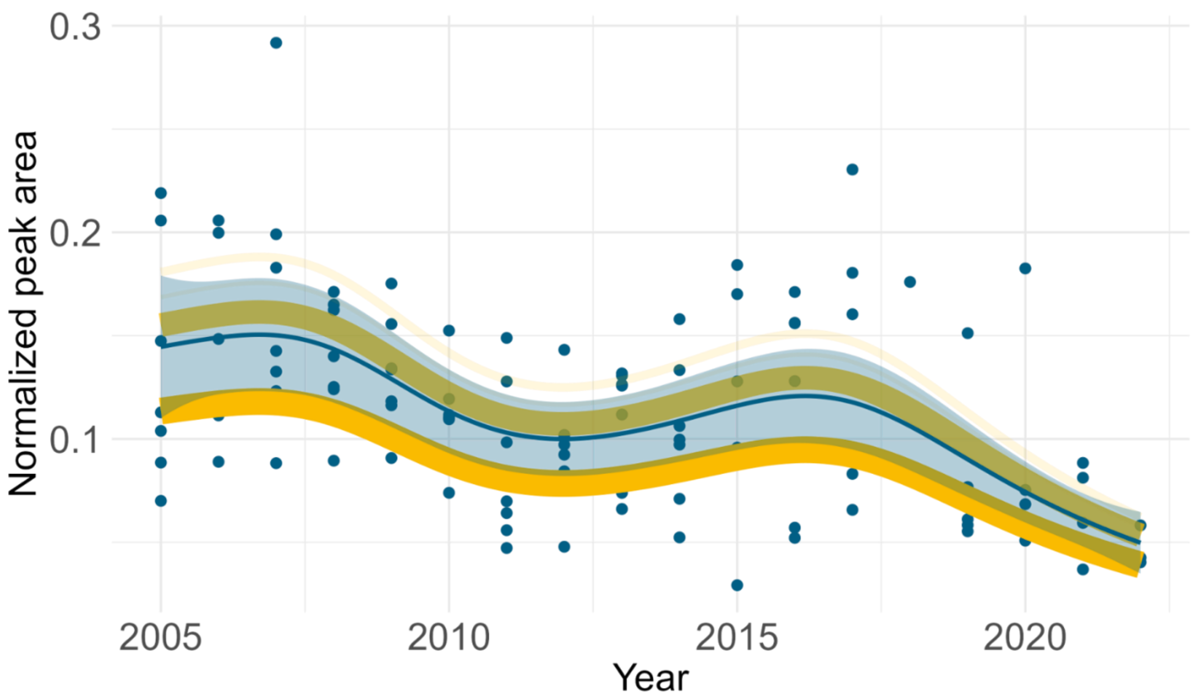

Figure S 2-8: Normalized peak areas of carbendazim in SPM samples from the Elbe (E1, E2, E4), Mulde (Mu), Rhine (R3, R4), and Saale (Sa) rivers. Blue points indicate individual annual measurements, while the lines represent the GAM Gamma fit (p(year) < 0.0001, slope = –0.01) with a weighted smooth and associated 95% confidence interval. Orange lines indicate river-specific contributions, with darker and thicker lines reflecting stronger influence.

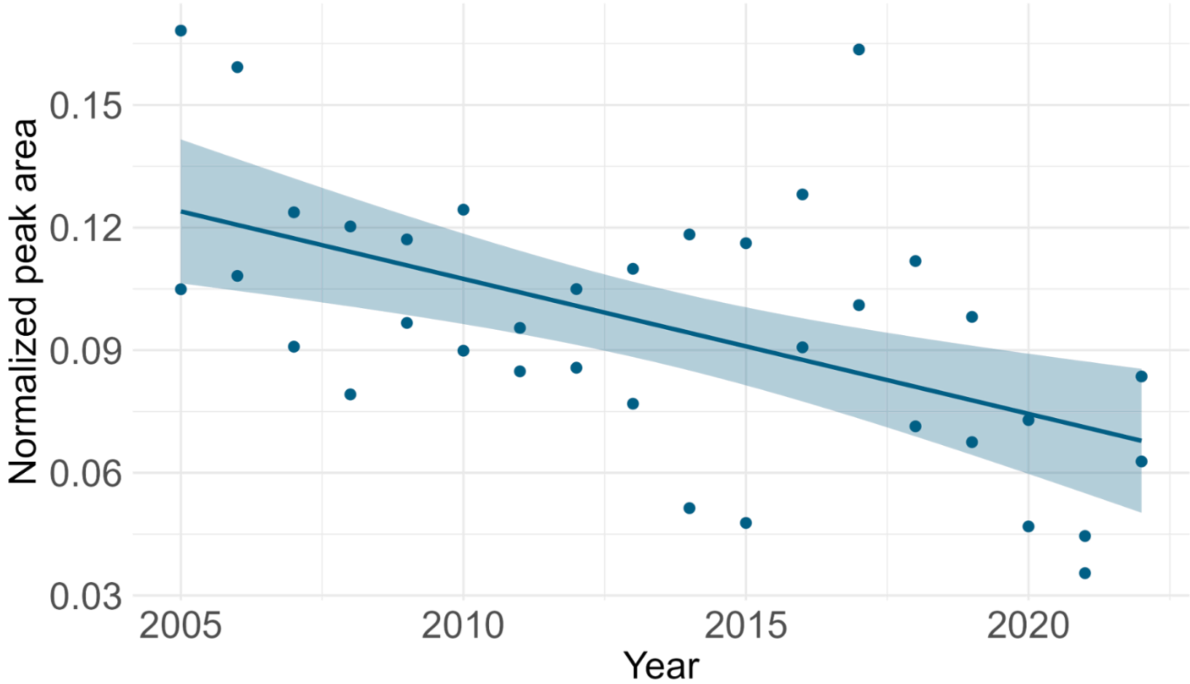

Figure S 2-9: Normalized peak areas of fludioxonil in SPM samples from the Rhine (R3, R4). Blue points represent individual annual samples. Blue line shows the overall GLM fit and 95% confidence interval (p(year) = 0.001, slope = 0.00, note y-axis dimension).

S-2.7. Chemical legislation – Human and Veterinary Medicinal Products

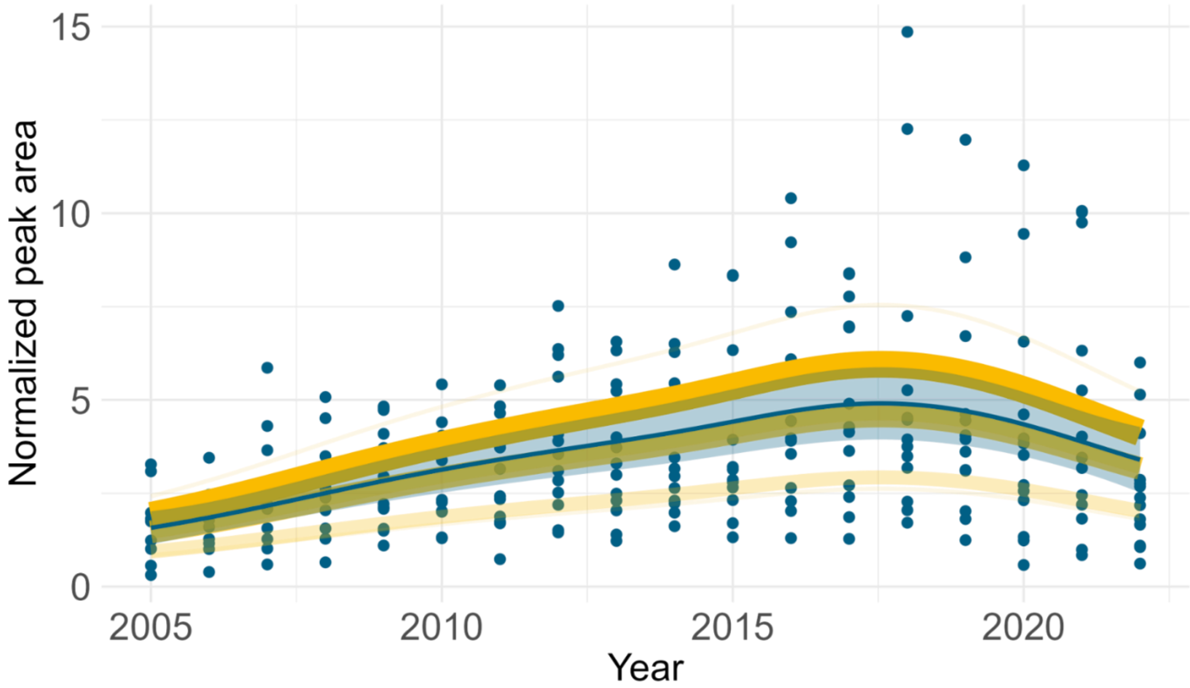

Figure S 2-10: Normalized peak areas of citalopram in SPM samples from the Elbe (E1, E2, E4, E5), Mulde (Mu), Rhine (R1, R3, R4), Saale (Sa), Saar (S1, S2), and Danube (D1, D3) rivers. Blue points represent individual annual samples. Blue line shows the overall GAM Gamma fit with a weighted smooth and 95% confidence interval (p(year) < 0.0001, slope = 0.11). Orange lines indicate river-specific contributions, with darker and thicker lines reflecting stronger influence.

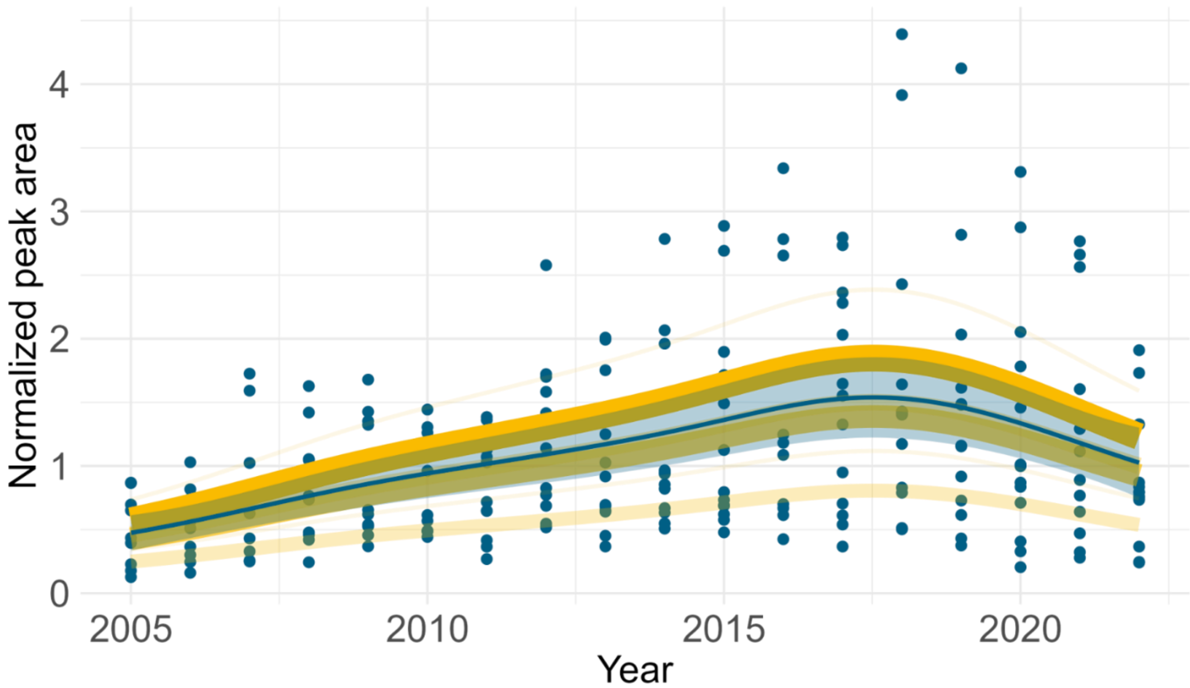

Figure S 2-11: Normalized peak areas of N-desmethyl-citalopram in SPM samples from the Elbe (E1, E2, E4, E5), Mulde (Mu), Rhine (R1, R3, R4), Saale (Sa), Saar (S1, S2), and Danube (D1, D3) rivers. Blue points represent individual annual samples. Blue line shows the overall GAM Gamma fit with a weighted smooth and 95% confidence interval (p(year) < 0.0001, slope = 0.03). Orange lines indicate river-specific contributions, with darker and thicker lines reflecting stronger influence.

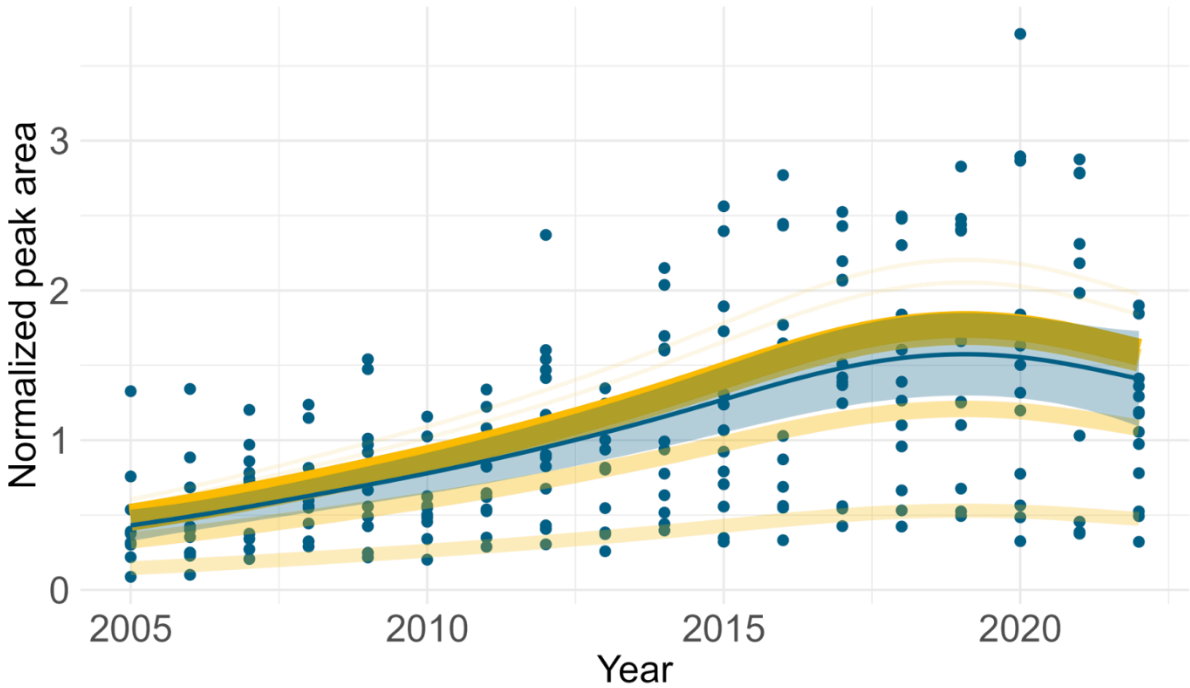

Figure S 2-12: Normalized peak areas of Venlafaxine in SPM samples from the Elbe (E1, E2, E4, E5), Mulde (Mu), Rhine (R1, R3, R4), Saale (Sa), Saar (S1, S2), and Danube (D1, D3) rivers. Blue points represent *individual annual samples. Blue line shows the overall GAM gamma fit with a weighted smooth and 95% confidence interval (p(year) < 0.0001, slope = 0.06). Orange lines indicate river-specific contributions, with darker and thicker lines reflecting stronger influence.*