# ENVIRONMENTAL MONITORING FOR MICROPOLLUTANTS USING HIGH-RESOLUTION MASS SPECTROMETRY AND DATA-DRIVEN METHODS

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Corey Michael George Carpenter

May 2019

# ENVIRONMENTAL MONITORING FOR MICROPOLLUTANTS USING HIGH-RESOLUTION MASS SPECTROMETRY AND DATA-DRIVEN METHODS

Corey Michael George Carpenter, Ph.D.

Cornell University 2019

Currently monitored contaminants represent only a fraction of the total chemicals present in natural water resources. There are over 100,000 chemicals used today and population growth has amplified global chemical manufacturing and production. The human and ecological exposome, a total measure of exposures over a lifetime, is poorly understood – particularly with respect to micropollutant exposure. Micropollutants are trace organic contaminants and represent a diverse set of chemicals such as pharmaceuticals, personal care products, pesticides, and industrial compounds. Many micropollutants were designed to be persistent and bioactive, and as a result, they can accumulate in the environment far away from their sources and their toxicological effects can be severe (*e.g.*, carcinogenic, interfere with endocrine systems, and cause anti-bacterial resistance). Micropollutants often occur at low concentrations in water resources ($ng \cdot L^{-1}$ to $\mu g \cdot L^{-1}$ range), which poses challenges for modern analytical technologies. The sources and spatiotemporal dynamics of micropollutants in the environment are also poorly understood, which presents challenges for developing regulations and mitigation strategies. The recent development of high-resolution mass spectrometry (HRMS) allows for the simultaneous detection of potentially thousands of known and unknown micropollutants by allowing extremely precise and accurate measurements at low concentrations. However, HRMS techniques have not often been utilized for environmental micropollutant monitoring, partly due to the lack of universal standard methods.

Additionally, covariates such as geospatial features and spatiotemporal environmental conditions are infrequently coupled with monitoring data to improve our understanding of the processes that control spatiotemporal micropollutant dynamics.

Three micropollutant monitoring studies were designed to more fully characterize and understand the surface water exposome by coupling broad micropollutant characterization afforded by HRMS with data-driven methods to further our understanding of micropollutant sources, fate, and transport. The first study incorporates spatially distributed sampling throughout a large watershed to discover links between the occurrence and concentrations of micropollutants, geospatial features of the watershed, and micropollutant sources. The second study incorporates highly resolved temporal sampling and utilizes micropollutant occurrence trend patterns to prioritize analytical data for identification of unknown micropollutants. The third study couples highly resolved micropollutant concentration trends with a set of environmental covariates to further our understanding of how environmental processes control micropollutant dynamics in surface waters as a means to predict peak events and inform intermittent sampling strategies. These environmental monitoring techniques and their results will aid future micropollutant monitoring campaigns to obtain more representative results and enable better management of micropollutants in surface water systems.

# BIOGRAPHICAL SKETCH

Throughout Corey Carpenter's academic career, he has strived to improve water quality knowledge through local investigations with global environmental implications. Corey Carpenter was born and raised in Boston, Massachusetts. He received a Bachelor's of Science degree in Environmental Engineering from Syracuse University (2014). While studying at Syracuse University, he worked with Dr. Charles Driscoll at the Center for Environmental Systems Engineering investigating how stormwater management practices can improve local water quality. He also completed an internship with Raymond International, where he traveled across the Middle East studying the means and ways of transporting potable water in arid regions. Later, he joined the Environmental Processes department in the School of Civil and Environmental Engineering at Cornell University to pursue his Masters and Doctorate degrees under the advisement of Dr. Damian E. Helbling. He completed his Masters of Science degree in Civil and Environmental Engineering with a thesis entitled: "Removal of micropollutants in biofilters: Hydrodynamic effects on biotransformation rates" (2016). During his academic career at Cornell University, he received the Graduate School Fellowship (2014) and the McMullen Fellowship (2015), and several awards including the Bernard Meyers Graduate Fellowship Award (2016), Honorable Mention from the National Science Foundation Graduate Research Fellowship Program (2016), Integral Consulting Inc. Research Award (2016), Conference Travel Grant Award (2016 and 2018), Certificate of Merit from the American Chemical Society (2016), Graduate Student Award in Environmental Chemistry from the American Chemical Society (2018), and had an article selected for the American Chemical Society Editors' Choice Award (2018).

Research conducted at Cornell University was disseminated in the following peer-reviewed publications:

(1) **Carpenter, C.M.G.**; Helbling, D.E. Removal of Micropollutants in Biofilters: Hydrodynamic Effects on Biofilm Assembly and Functioning. *Water Research.* **2017**, *120*, 211–221. https://doi.org/10.1016/j.watres.2017.04.071. (Master's Thesis).

(2) **Carpenter, C.M.G.**; Helbling, D.E. Widespread Micropollutant Monitoring in the Hudson River Estuary Reveals Spatiotemporal Micropollutant Clusters and Their Sources. *Environmental Science & Technology.* **2018**, *52* (11), 6187–6196. https://doi.org/10.1021/acs.est.8b00945. **(Chapter 2)**.

(3) **Carpenter, C.M.G.**; Wong, L.Y.J.; Johnson, C.A.; Helbling, D.E. Fall Creek Monitoring Station: Highly Resolved Temporal Sampling to Prioritize the Identification of Nontarget Micropollutants in a Small Stream. *Environmental Science & Technology.* **2019**, *53* (1), 77–87. https://doi.org/10.1021/acs.est.8b05320. Selected as American Chemical Society Editors' Choice on 12 December 2018. **(Chapter 3)**.

(4) Gao, H.; LaVergne, J.M.; **Carpenter, C.M.G.**; Desai, R.; Zhang, X.; Gray, K.; Helbling, D.E.; and Wells, G. Exploring Co-Occurrence Patterns between Organic Micropollutants and Bacterial Community Structure in a Mixed-Use Watershed. *Environmental Science: Processses & Impacts.* **2019**, *Advance Article*. https://doi.org/10.1039/c8em00588e.

(5) **Carpenter, C.M.G.**; Wong, L.Y.J.; Gutema, D.L.; Helbling, D.E. Fall Creek Monitoring Station: Using Environmental Covariates to Predict Micropollutant Dynamics and Peak Events in Surface Water Systems. *Submitted for publication.* **(Chapter 4)**.

(6) Lin, Y.; Jiang, T.; Rivera, M.S.; Cotto, I.; **Carpenter, C.M.G.**; Rich, S.L,; Larese-Casanova, P.; Giese, R.W.; Helbling, D.E.; Pinto, A.; Padilla, I.Y.; Rosario-Pabón, Z.; Alshawabkeh, A.N.; Gu, A.Z. Integrating Chemical Analysis and a High-Throughput Quantitative Toxicogenomics Approach to Assess the Impact of Hurricane Maria on Drinking Water Quality in Puerto Rico. *In preparation.*

(7) Tarpeh, W.A.; Du Y.; **Carpenter, C.M.G.**; Helbling D.E.; Aga D.; Love, N.G.; Wigginton, K.R. A Unit Process Approach to Screening Pharmaceuticals during Urine Treatment. *In preparation.*

# ACKNOWLEDGMENTS

First and foremost, I would like to thank my major advisor, Dr. Damian E. Helbling, for his continued support throughout my time at Cornell University. His guidance and mentorship were instrumental to the successful completion of this dissertation. I came out of every meeting feeling invigorated and ready to take on the next steps towards our research goals. I would also like to thank Dr. Todd Walter and Dr. David Rossiter for serving as minor advisors on my committee and for the valuable feedback and discussions. I am thankful for the rest of the Helbling Research Group (HRG) for their friendship and support. I want to acknowledge undergraduate students Catie, Danyeh, and Jacqueline for their assistance with the Fall Creek Monitoring Station – thank you for all your hard work! Past and present HRG members have helped me become a better researcher, both inside and outside the laboratory. I am also grateful for the support that came from my friends and family. I would like to specifically thank my wife, Sarah Wilkinson, for her support throughout my doctoral studies – your encouragement and love helped me more than I can possibly describe.

**GO BIG RED!**

# TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# LIST OF ABBREVIATIONS

| | |
|---|---|
| AICc | corrected Akaike information criterion |
| ANOVA | anlaysis of variance |
| AUC | area under curve |
| BMP | best management practice |
| CCL | contaminant candidate list |
| CWFP | Cornell Water Filtration Plant |
| EAR | exposure-activity ratio |
| EDA | effect-directed analysis |
| EPA | Environmental Protection Agency |
| EU | European Union |
| FC | Fall Creek |
| FCMS | Fall Creek Monitoring Station |
| GC | gas chromatography |
| GLS | generalized least squares |
| HCA | hierarchical clustering analysis |
| HPLC | high performance liquid chromatography |
| HRE | Hudson River Estuary |
| HRMS | high-resolution mass spectrometry |
| ILIS | isotope labeled internal standards |
| logP | octanol-water partition coefficient |
| LOQ | limit of quantification |
| LVI | large volume injection |
| *m/z* | mass-to-charge ratio |
| MoNA | MassBank of North America |
| MPP | micropollutant profile |
| MS | mass spectrometry |
| $MS^1$ | mass spectra |
| MS/MS | tandem mass spectrometry |
| $MS^2$ | fragmentation mass spectra |
| NTA | nontarget analysis |
| PFAS | perfluoroalkyl substance |
| ROC | receiver operating characteristic |
| RT | retention time |
| SPE | solid phase extraction |
| STP | sewage treatment plant |
| TP | transformation product |
| Tukey-HSD | Tukey’s honest signficant difference |
| U.S. | United States |
| USGS | United States Geological Survey |
| VIF | variance inflation factor |
| WRS | Wilcoxon rank sum |

# CHAPTER 1 – Background

## 1.1 Micropollutants

Organic chemicals are used daily in a variety of ways to improve the lives of the global population and contribute to ubiquitous facets of everyday life; however, it has been documented globally that anthropogenic trace organic chemicals, collectively known as micropollutants, are entering and accumulating in our natural water resources.[1] We use pharmaceuticals and personal care products to protect human and animal health, antimicrobials to defend against the spread of disease, and pesticides to increase agricultural crop yields. These and other micropollutants enter the aquatic environment through point sources (*e.g.*, sewage treatment plants (STPs) and industrial discharge) and nonpoint sources (*e.g.*, runoff from agriculture and septic effluents); although, many other sources remain unknown.[2–4] Once in the environment, micropollutants can undergo abiotic and biotic transformation processes yielding many transformation products, some of which are more mobile and/or toxic than the parent compound.[5,6] Micropollutants have been detected throughout the urban water cycle, from STP influent and effluent to surface water resources to drinking water treatment plant influent and effluent.[7–10]

Major concerns over human and ecological micropollutant exposure prompted *introduction of novel entities* to be included as one of the nine planetary boundaries, which describe a framework that regulates the stability of the Earth system.[11] Population growth has amplified global chemical manufacturing and production. Numerous novel entities such as anthropogenic chemicals are entering the environment for the first time, which is of global concern when these chemicals are persistent, mobile across widespread areas, and have the potential for adverse effects. The *introduction of novel entities* planetary boundary has not yet been quantified and

remains highly uncertain, partly due to the lack of broad environmental monitoring data for micropollutants. Even though micropollutants are typically found at low concentrations ($ng \cdot L^{-1}$ to $\mu g \cdot L^{-1}$ range), they can have adverse effects on the environment and human health.[12–14] It has been shown that at environmentally relevant concentrations, some micropollutants can be carcinogenic, can interfere with human and animal endocrine systems, or cause anti-bacterial resistance, but the effects of many others are still unknown.[15–19] For example, it is widely documented that exposure to micropollutants such as the herbicide atrazine and synthetic estrogen can cause the feminization of frogs and fish in contaminated aquatic systems.[19,20] Urban stormwater mortality syndrome in coho salmon has been linked to specific micropolltuants.[21] Human exposure to an emerging class of micropollutants called perfluoroalkyl substances (PFASs) in drinking water resources has been linked to increased cancer incidence.[22]

Micropollutant contamination in the environment and in drinking water resources is not broadly regulated due to the lack of monitoring data and risk assessments. In the United States (U.S.), the Environmental Protection Agency (EPA) maintains a contaminant candidate list (CCL) of potential chemicals that may be regulated in the future.[23] The CCL contains a growing number of micropollutants; however, the Safe Drinking Water Act only regulates 20 pesticides and 1 transformation product through the National Primary Drinking Water Regulations.[24] There are over 100,000 chemicals registered for use in the U.S. through the EPA's Toxic Substances Control Act and Federal Insecticide, Fungicide, and Rodenticide Act, and approved by the U.S. Food & Drug Administration.[25–27] These registries do not entirely account for the possibility of transformation products; each registered chemical may have multiple transformation products, which dramatically amplifies the number of chemicals added to the environment by anthropogenic activities.[28] Regulations in the U.S. are based on monitoring programs and health risk assessments;

however, routine environmental monitoring for micropollutants is uncommon and the sources and spatiotemporal dynamics of micropollutants in the environment remain poorly understood, which presents challenges for developing regulations and mitigation strategies to protect human health and the environment against the adverse effects of micropollutant exposure. There are far too many chemicals to systematically regulate and many chemicals are present in the environment that we may not be aware of yet. In the European Union (EU), regulations are based on the precautionary principle, which states that the introduction of novel chemicals to the environment that have unknown effects should be limited.[29] Specifically, the EU regulates pesticides in drinking water under the Drinking Water Directive using a maximum allowable concentration of 0.1 $\mu g \cdot L^{-1}$ for any individual chemical as a surrogate zero concentration and 0.5 $\mu g \cdot L^{-1}$ for cumulative pesticide concentration. Although, this method of applying a single limit to all pesticides does not account for available environmental monitoring or toxicological data. Improved micropollutant monitoring approaches are needed to contribute broad micropollutant occurrence data that can inform risk assessments and environmental regulators.

As an example of the consequences of the current lack of broad micropollutant occurrence data, we can examine the U.S. regulation history of PFASs. Two major PFASs, perfluorooctanoic acid (PFOA) and perfluorooctanesulfonic acid (PFOS), have been used in many consumer products, industrial applications, and in firefighting foams since the mid-twentieth century. However, concerns over PFASs in water resources led to the phase-out of U.S. production of PFOS in 2002, and several global companies' production of PFOA in 2006. However, new PFASs emerged – in 2015, over 40 unique PFASs were detected in global drinking waters with the potential for 100s of others.[30] In 2016, the EPA issued a drinking water health advisory for the combined concentrations of PFASs (PFOS and PFOA) at 70 $ng \cdot L^{-1}$, which exemplifies the

potential health risks (developmental effects, cancer, and others) of micropollutants at trace concentrations. PFASs have since been detected in adult blood plasma and global drinking waters, indicating high human exposure.[30,31] The only health advisory for PFASs was issued over ten years after the phase-outs, only includes two chemicals, and there are no current enforceable regulations despite known human health risks.

### 1.2 Monitoring approaches

Broad micropollutant monitoring studies in natural water systems have only recently been implemented and the state-of-the-science approaches have largely remained unchanged. The first major national reconnaissance for micropollutants in the U.S. was conducted by the U.S. Geological Survey (USGS) in 2002 and used a targeted monitoring approach to characterize a predefined set of micropollutants in single samples from multiple streams.[3] Target analyses require *a priori* information about the micropollutants of interest by first analyzing pure reference standards that can then be compared with field samples to identify and quantify those specific micropollutants. The breadth of micropollutants examined is usually quite limited due to the high cost of acquiring pure reference standards and time-intensive post-acquisition data processing methods. More recently in 2017, the USGS conducted another national reconnaissance for micropollutants using a similar approach to characterize micropollutants in single samples from multiple streams.[32] Narrow approaches will often underrepresent true micropollutant contamination and associated risks by either not targeting relevant micropollutants or having insufficient sample resolution.[33]

Target analyses are the most frequently used approach in micropollutant monitoring studies but remain extremely variable in their design in several ways. First, these studies monitor different

numbers of micropollutants depending on the analytical technologies employed, from as few as one to as many as 893 target micropollutants.[32] Second, conventional micropollutant monitoring approaches vary widely in spatiotemporal sample resolution: from low temporal resolution and high spatial resolution – 38 one-time grab samples from streams across the U.S.,[32] to high temporal resolution with low spatial resolution – daily samples at one station for over two years,[34] and medium temporal resolution with medium spatial resolution – seven monthly grab samples from four sites.[35] High spatial resolution sampling can lead to key insights about specific sources of micropollutants, while high temporal resolution sampling can capture and describe the dynamic micropollutant profiles which change with seasons, use patterns, and environmental conditions.[36,37] Lastly, the type of sampling method varies and can influence micropollutant characterization results.[38] Grab samples can only give a snapshot of micropollutant occurrence at a specific site and time, but are more easily obtained. Flow or time-weighted composite samples require specialized sampling equipment, but can give a better overall representation of micropollutant occurrence and concentrations. Variable targeted approaches are often not comprehensive enough and more holistic micropollutant monitoring approaches are needed to provide insights into micropollutant sources and spatiotemporal dynamics in a way that can inform best management practices (BMPs) for micropollutant management at watershed scales.

### 1.3 Data-driven methods

Once micropollutant occurrence data has been acquired, various data-driven methods have been employed to expand our knowledge about micropollutant sources and spatiotemporal dynamics. Data-driven unsupervised methods are directed by the data without user bias and are used to extract insights directly from the data itself. Data-driven methods such as hierarchical

clustering analysis (HCA) have been used to identify similar groups of micropollutants that co-occur in the environment.[35,39,40] HCA groups sets of observations based on their similarity; sets of observations that are clustered closely are more similar than those that are clustered farther away. Sets of observations can include micropollutant occurrence data from different sites and/or at different times. With spatially distributed sampling, HCA groups sites with similar micropollutant contamination patterns and when combined with geospatial data (watershed features such as land cover and locations of point sources) can lead to source attribution.[36] With high temporal resolution sampling, HCA groups micropollutants with similar temporal trends and therefore similar fate and transport processes.[41] Data-driven methods have also been employed for model selection to find the parameters that can best predict micropollutant degradation, adsorption, and temporal dynamics.[42–44] Model selection techniques can lead to a better understanding of the drivers and mechanisms underlying basic micropollutant processes such as transformation, adsorption, and transport.

### 1.4 Micropollutant characterization using mass spectrometry

Sample pretreatment and preparation including cleanup and concentration methods are often required prior to mass spectrometry (MS) analyses to isolate and concentrate analytes from environmental water samples. First, solid particulates are removed through centrifugation or filtration. Then, sample enrichment methods are performed, often with subsequent evaporation to concentrate the sample. Since micropollutants regularly occur at low concentrations in the environment, sample enrichment methods are typically necessary; solid phase extraction (SPE) is commonly used. SPE isolates micropollutants based on their physiochemical properties onto solid materials and allows other substances to pass through, thus providing matrix cleanup. The isolated

micropollutants can then be eluted by an organic solvent, which is subsequently evaporated, and the sample is finally reconstituted into water. Typical SPE procedures for micropollutant characterization can concentrate a water sample by more than 2 – 3 orders of magnitude. However, these methods can often be time-intensive and cost prohibitive when working with a large number of samples.

Following sample preparation, gas chromatography (GC) or high performance liquid chromatography (HPLC) coupled with MS is used for the characterization of micropollutants. These methods can be used to identify and quantify compounds in complex mixtures including surface water,[45] groundwater,[46] STP effluent,[47] hydraulic fracturing (fracking) fluid,[48] food,[49] blood,[50] and others. Polar to semi-polar organic contaminants are generally more hydrophilic and partition more favorably towards the water phase in the environment and therefore are more amenable to HPLC analysis. Non-polar organic contaminants represent about 20 – 30% of the total organic carbon in water and are detectable by GC, while polar organic contaminants represent the remaining 70 – 80% and are detectable by HPLC.[51] Aquatic micropollutant characterization analytical methods often utilize reversed phase HPLC to separate analytes based on their polarity using a hydrophobic solid column and a liquid mobile phase consisting of an organic and a water phase. In multiresidue analyses, the mobile phase is often delivered to the column at a gradient to permit optimal separation of multiple analytes. The separation of analytes ensures that the mass spectrometer is not overloaded at any given time and can separate micropollutants that may otherwise be indistinguishable (*e.g.*, isomers). The time at which a compound is measured in the mass spectrometer along the analytical gradient is known as the retention time (RT).

Mass spectrometry can simultaneously generate data on many compounds. After a sample is separated using chromatography, it enters an ion source (*e.g.*, electrospray ionization) where it

is vaporized and acquires either a positive or negative charge. During positive mode ionization, the parent compound [M] typically forms an adduct with a variety of positively charged species including protons $[M+H]^+$ and sodium ions $[M+Na]^+$. In negative mode ionization, the parent compound typically undergoes deprotonation to form $[M-H]^-$. After ionization, the sample enters the mass spectrometer where the charged species are separated and/or filtered and detected as a mass-to-charge ratio (*m/z*) and intensity. The *m/z* is determined by the mass of the charged parent compound species divided by the charge and the intensity is a discrete measure of how many particular ions are present in the sample and scales directly with concentration.

### 1.5 Advantages of high-resolution mass spectrometry

In recent years, the development of accurate-mass high-resolution mass spectrometry (HRMS) has introduced new opportunities for environmental micropollutant monitoring.[52] As described in **Figure 1.1**, HRMS technologies can measure the exact masses of compounds out to five decimal places with high accuracy, while conventional mass spectrometry (*e.g.*, triple

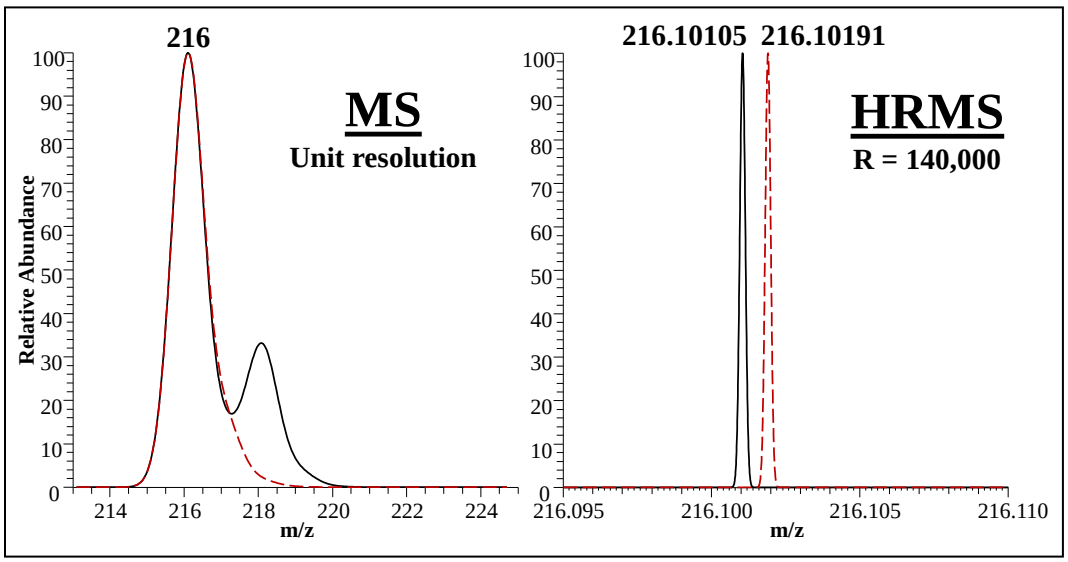



**Figure 1.1:** Mass spectra acquired with mass spectrometry (left) and high-resolution mass spectrometry (right) for atrazine (solid black line) and furcarbanil (dashed red line).

quadrupole) with unit resolution can only measure accurately at the integer level. In this example, HRMS is able to distinguish between two herbicides (atrazine and furcarbanil) that have a mass difference of only 0.00086 amu. Environmental monitoring methods can take advantage of HRMS because, with improved resolving power (R = 5,000 – 240,000 at 200 *m/z*) and mass accuracy ($\Delta$m = <1 – 5 ppm), HRMS is capable of simultaneously detecting potentially thousands of compounds by collecting data for each detectable ion in the sample – referred to as full-scan MS. Each compound detected by HRMS is described by an *m/z*, an isotopic signature, and a RT. An example of raw data acquired from quadrupole-orbitrap HRMS is provided in **Figure 1.2**. The monoisotopic exact mass is defined by the chemical formula of the compound; it is simply a

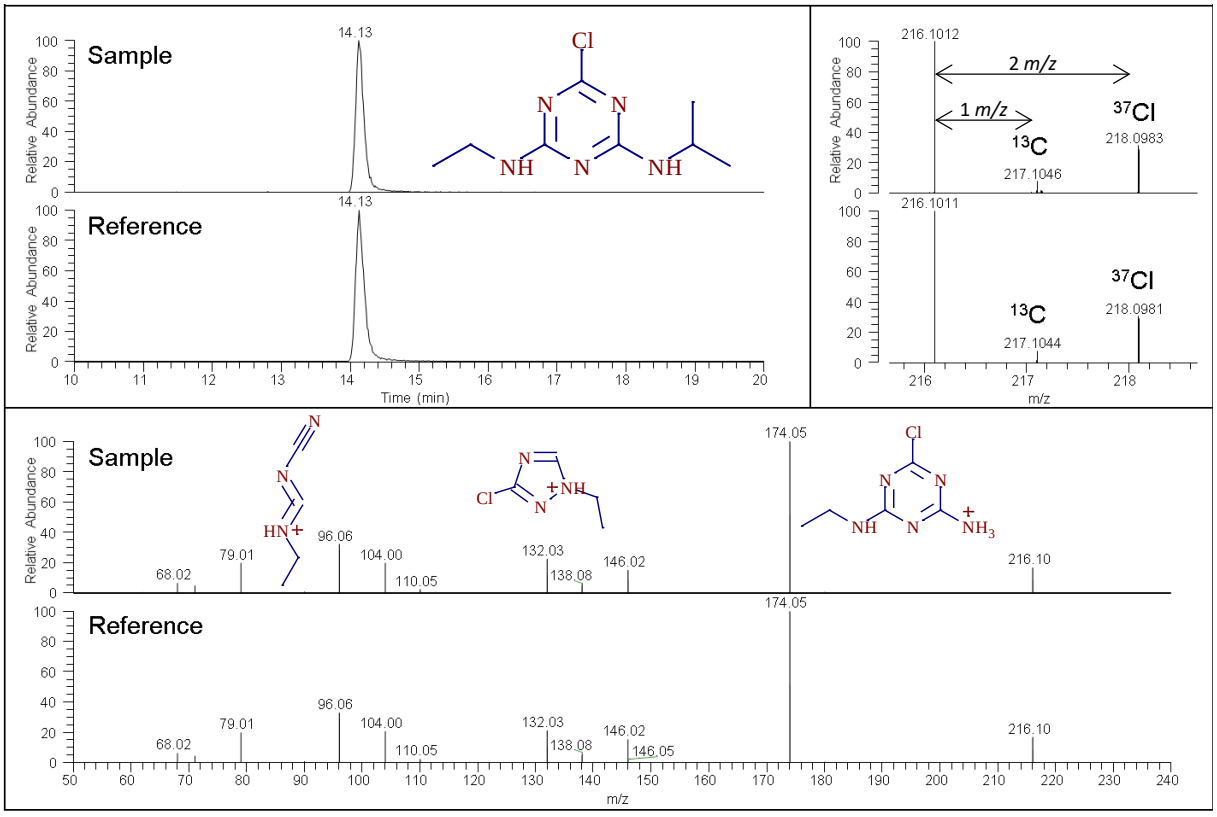



**Figure 1.2:** Example of raw mass spectrometry data – confirmation of atrazine in river water. Atrazine ($C_8H_{14}ClN_5$) was confirmed using an authentic reference standard which matched the RT of 14.13 min, the MS spectra (*m/z* =216.1012 for $[M+H]^+$, $\Delta$m = 0.69 ppm), the theoretical abundance (8%) of the $^{13}C$ monoisotopic mass, the theoretical abundance (32%) of the $^{37}Cl$ monoisotopic mass, and the MS/MS fragments with *m/z* = 68.02, 96.06, 104.00, 110.05, 132.03, 138.08, 146.02, and 174.05. Structures of underlined fragments are shown.

summation of the monoisotopic exact masses (in amu) of each atom that makes up a compound. The isotopic signature includes additional mass spectra for the natural abundances of isotopes of certain atoms such as 1% $^{13}C$, 4.5% $^{34}S$, and 32% $^{37}Cl$ (top right panel of **Figure 1.2**). Ions belonging to the same compound, such as adducts and isotopes, will elute at the same RT and have similar peak shapes. A chromatographic peak refers to the intensity pattern surrounding an eluting ion, which is usually a Gaussian curve centered at the RT (top left panel of **Figure 1.2**). Finally, with hybrid MS systems, fragmentation data can also be acquired using a collision cell, which bombards the molecules with energy to break up the chemical structure into smaller fragments and records the accurate masses of the MS fragments as fragmentation spectra (MS/MS, bottom panel of **Figure 1.2**). Fragments can be used to differentiate co-eluting isomeric compounds and to confirm the presence of compounds in field samples. A micropollutant is unambiguously confirmed if it is detected at the same *m/z* and RT as the reference standard, and has matching MS/MS.

### 1.6 Nontarget analysis

Nontarget analysis (NTA) is a powerful micropollutant characterization technique that exploits full-scan HRMS data. NTAs use automatic peak-picking software (*e.g.*, enviMass[53]) to identify fully-resolved chromatographic peaks in HRMS data, resulting in a peak list for each sample. Each peak represents an individual compound or is related to a compound through isotopologues (isotope mass spectra), adducts (multiple charged species), or in-source fragments (caused by the ionization source). Unlike highly specific target analyses, NTAs are nonspecific and do not require *a priori* information about micropollutants of interest. While target screenings are still the routine approach for micropollutant monitoring, more researchers are moving towards

NTAs to comprehensively characterize micropollutants in a variety of systems including sediments,[41] animal fat,[54] dust,[55] STP effluent,[56] surface water,[57] and others.[34,58] However several challenges still limit NTA including the need for improved data reduction and prioritization methods. Since NTA utilizes full-scan MS data, many data reduction techniques and prioritization methods have been developed to reduce the size of the resultant peak lists and to sort these lists by relevance prior to nontarget structure elucidation. For example, typical NTA data reduction methods include blank subtraction to remove peaks detected in both blanks and samples, replicate filters to remove peaks that are not ubiquitously present in all replicates, and grouping algorithms (componentization) to combine peaks that are related to the same compound through adducts and isotopes into components.[34] Past data-driven prioritization approaches focused on peaks with high intensities,[56] high frequencies of detection,[59] homologous series,[60] and spatiotemporal profiles.[37]

NTA ultimately relies on precise measurement of exact masses, isotopic signatures, and fragmentation spectra to make accurate chemical formula predictions and structure elucidations. Nontarget peaks are assigned a confirmation level based on the certainty of the prediction as described in **Table 1.1**.[61] Once a peak list is prioritized, the analyst can examine the highest ranked peaks to propose structures of unknown micropollutants. Chemical formulas are predicted through the exact mass of the chromatographic peak and the isotopic signature is used to constrain the atomic boundaries of the prediction. For example, if the characteristic $^{37}Cl$ isotope is not present in the mass spectra, then it can be assumed that the compound does not contain any Cl atoms. Additionally, if the $^{13}C$ isotope is at 10% relative abundance when compared to the monoisotopic intensity, then it can be assumed that about 10 C atoms are in the compound. Then, the predicted formulas can be searched for in chemical databases (*e.g.*, ChemSpider[62] or PubChem[63]) to obtain a list of potential chemical structures. The chemical structures can be examined and confirmed

through MS/MS spectra matching with either predicted fragments (*e.g.*, MetFrag[64]) or library fragments (*e.g.*, MassBank[65]). The predictions can be inspected further by matching spectra with chemical databases and unambiguously confirmed with authentic reference standards.

**Table 1.1:** Identification confidence levels for high-resolution mass spectrometry analyses adapted from Schymanski et al., 2014.[61]

| Confirmation Level[a] | | Confirmation method |
|---|---|---|
| **Level 1** | Confirmed structure | Matching RT (reference standard) |
| **Level 1L** | Confirmed structure | Matching library RT (in-house library) |
| **Level 2L** | Probable structure | Matching library MS/MS (MassBank) |
| **Level 2P** | Probable structure | Matching predicted MS/MS (MetFrag) |
| **Level 3** | Tentative candidate(s) | MS/MS evidence (MetFrag) |
| **Level 4** | Unequivocal molecular formula | Distinct isotopic signature |
| **Level 5** | Exact mass of interest | Prioritized peak |

[a] Confirmation levels are accumulative and sequential.

### 1.7 Research objectives

Over the past two decades, research focused on the impacts of chemical pollution in surface waters has slowly broadened from conventional priority pollutants to include micropollutants.[32,46,66,67] However, we are still limited in our ability to comprehensively characterize micropollutant contamination in a way that can lead to BMPs for micropollutant management at a watershed scale. Comprehensive monitoring campaigns are essential for developing risk assessments and mitigation strategies. Broadly characterizing micropollutant occurrence and better understanding the sources, fate, and transport of micropollutants are necessary to make educated recommendations to stakeholders for water quality improvement. Current typical micropollutant monitoring studies screen relatively few samples for relatively few micropollutants. This type of monitoring approach cannot lead to comprehensive assessments of micropollutant occurrence, concentration, toxicity, or sources because the resulting datasets lack spatial resolution within individual watersheds, temporal resolution, and broad micropollutant

characterization. Additionally, covariates such as geospatial features and spatiotemporal environmental conditions are infrequently coupled with micropollutant monitoring data to improve our understanding of micropollutant sources and the processes that control spatiotemporal dynamics. With relatively small datasets, it is difficult to exploit data-driven methods without adding user and sample bias.

This research aims to address these issues to improve environmental micropollutant monitoring approaches. In recent years, advances in MS technologies, improved abilities to collect environmental covariates, and enhanced computational data-driven tools have provided the unique opportunity to investigate these knowledge gaps. This research strives to develop micropollutant monitoring campaigns capable of comprehensively characterizing micropollutant occurrence and spatiotemporal dynamics, while simultaneously linking the resultant datasets with environmental covariates in ways that can generalize or contextualize the micropollutant characterization results. The overarching goal of this research is to improve our micropollutant monitoring capabilities to more comprehensively characterize micropollutant contamination and to further our understanding of the sources, fate, and transport of micropollutants in surface waters by coupling broad micropollutant monitoring afforded by HRMS with data-driven methods. The following three research projects were employed to meet this goal.

Project A, entitled *Widespread Micropollutant Monitoring in the Hudson River Estuary Reveals Spatiotemporal Micropollutant Clusters and Their Sources*, explores a spatially distributed sampling campaign throughout a large watershed at relatively infrequent times over two years to discover links between the occurrence and concentrations of micropollutants, geospatial features of the watershed, and micropollutant sources. By comprehensively characterizing micropollutant contamination throughout the entire watershed, groups of

micropollutants with similar spatiotemporal occurrence and concentration patterns are identified using HCA. The micropollutant clusters discovered in this study can be used to characterize micropollutant occurrence and sources in surface water systems around the world. The monitoring campaign also identifies specific sub-watersheds that should be prioritized for implementation of micropollutant mitigation strategies.

Project B, entitled *Fall Creek Monitoring Station: Highly Resolved Temporal Sampling to Prioritize the Identification of Nontarget Micropollutants in a Small Stream*, incorporates highly resolved temporal sampling and uses micropollutant occurrence trend patterns to prioritize analytical data for identification of unknown micropollutants and to comprehensively characterize micropollutant contamination. For the first time, highly resolved temporal profiles of target micropollutants and nontarget MS features are generated using a continuous monitoring station on a small stream. The highly resolved temporal profiles are used to develop a novel data-driven prioritization technique based on HCA. The nontarget workflow is publicly available and applicable to HRMS data acquired from any type of sample. The highly resolved temporal data reveals the temporal dynamics of micropollutant occurrence to gain fundamental insights on contaminant sources, fate, and transport phenomena.

Project C, entitled *Fall Creek Monitoring Station: Using Environmental Covariates to Predict Micropollutant Dynamics and Peak Events in Surface Water Systems*, couples highly resolved micropollutant concentration trends with a set of environmental covariates to further our understanding of how environmental processes control micropollutant dynamics in surface waters as a means to predict peak events and inform intermittent sampling strategies. Environmental covariates that are predictive of the temporal concentration profiles and peak events of representative micropollutants and cumulative metrics of overall micropollutant contamination are

identified using data-driven model selection techniques. These predictors are used to develop intermittent sampling strategies aimed at capturing peak events and provide a more accurate representation of chemical risk. Finally, sampling triggers are provided as percentile values of predictive environmental covariates that can be used to inform water quality sampling that targets peak events in other watersheds and obtain representative micropollutant datasets that can more accurately assess micropollutant contamination and environmental risk in global surface waters.

# CHAPTER 2 – Widespread Micropollutant Monitoring in the Hudson River Estuary Reveals Spatiotemporal Micropollutant Clusters and Their Sources[a]


## Abstract

The objective of this study was to identify sources of micropollutants in the Hudson River Estuary (HRE). We collected 127 grab samples at seventeen sites along the HRE over two years and screened for up to 200 micropollutants. We quantified 168 of the micropollutants in at least one of the samples. Atrazine, gabapentin, metolachlor, and sucralose were measured in every sample. We used data-driven unsupervised methods to cluster the micropollutants based on their spatiotemporal occurrence and normalized-concentration patterns. Three major clusters of micropollutants were identified: ubiquitous and mixed-use (*core micropollutants*); sourced from sewage treatment plant outfalls (*STP micropollutants*); and derived from diffuse upstream sources (*diffuse micropollutants*). Each of these clusters was further refined into sub-clusters that were linked to specific sources based on relationships identified through geospatial analysis of watershed features. Evaluation of cumulative loadings of each sub-cluster revealed that the Mohawk River and Rondout Creek are major contributors of most *core micropollutants* and *STP micropollutants* and the upper HRE is a major contributor of *diffuse micropollutants*. These data provide the first comprehensive evaluation of micropollutants in the HRE and define distinct spatiotemporal micropollutant clusters that are linked to sources and conserved across surface water systems around the world.



[a] Reproduced with permission from *Environmental Science & Technology*, 2018, *52* (11), 6187–6196. DOI: 10.1021/acs.est.8b00945. Copyright 2018 American Chemical Society.

## 2.1 Introduction

Data from monitoring studies have routinely confirmed the occurrence of 100s of organic micropollutants in surface water systems around the world.[68] The main targets of monitoring studies have been pharmaceuticals,[69] personal care products,[47] illicit drugs,[70] pesticides,[33] industrial chemicals,[71] or other anthropogenic organic chemicals that have known or putative toxic effects on aquatic ecosystems or exposed human populations.[17,66,72] The potential sources of micropollutants are varied, with much attention focused on sewage treatment plant (STP) outfalls,[47] combined sewer overflows,[73] industrial discharges,[57] stormwater outfalls,[74] and diffuse runoff from agricultural and urban landscapes,[75] while many other potential sources are being explored.[76]

Recently, long-term monitoring data characterizing micropollutant occurrence at the watershed scale has been used to identify key insights into sources of micropollutants. For example, mass balance and multivariate analyses revealed three types of micropollutant sources in a Minnesota River including diffuse runoff, STP outfalls, and mixed pathways (diffuse runoff and STP outfalls).[35] Long-term longitudinal sampling along the Rhine River was used to identify several previously unknown sources of micropollutants, particularly from tributaries and industrial sources.[37] A geospatial analysis of poly- and perfluoroalkyl substances (PFASs) revealed that PFASs were found at higher concentrations in more urban areas and different types of PFASs were associated with different point sources such as airports, textile mills, and metal smelting.[36] Lastly, the U.S. Geological Survey (USGS) conducted a national-scale micropollutant monitoring survey and used a statistical approach to reveal significant relationships between contaminant summary statistics and wastewater discharge and urban development.[32] These examples demonstrate

powerful ways in which geospatial data can be combined with micropollutant occurrence data to improve our fundamental understanding of micropollutant sources.

The primary goal of this research was to assess the relative contributions of various sources of micropollutants in the Hudson River Estuary (HRE). The HRE provides drinking water to more than 100,000 people as a surface water source and is an important waterway for recreational and commercial activities. A recent study surveyed the occurrence of 16 pharmaceutical compounds in the HRE,[77] but no previous study has combined a comprehensive micropollutant screening with geospatial analyses to identify the relative contributions of various sources of micropollutants in the HRE. We hypothesized that groups of micropollutants would cluster together based on their spatiotemporal occurrence or concentration patterns, and that those clusters would associate with specific upstream sources. To test this hypothesis, we collected grab samples at seventeen sites along the HRE during the 2016 and 2017 recreational seasons (May – October). Samples were analyzed to quantify the occurrence of up to 200 micropollutants identified in surface waters around the world. We used ArcGIS to develop maps of the watershed that include geospatial references for likely micropollutant sources. We used data-driven unsupervised methods to explore the complexity of micropollutant occurrence, including hierarchical clustering to identify groups of micropollutants with similar spatiotemporal occurrence and normalized-concentration patterns. We were able to categorize the resulting micropollutant clusters based on their likely sources, link the clusters to various geospatial features, and assess the relative contributions of specific sources and tributaries to micropollutant occurrence in the HRE. We finally used a statistical approach to discover a contamination event and identify micropollutants that are suitable indicators of overall micropollutant occurrence and concentrations.

## 2.2 Material and Methods

### *2.2.1 Study area*

The HRE catchment area is a large mixed-use watershed located in eastern New York State with an area of approximately 34,300 $km^2$ and a population of over 2.5 million. A map of the study area, the locations of seventeen sampling sites, and a delineation of tributary watersheds is provided in **Figure 2.1**. Samples were collected from sites between the Mohawk River and the Tappan Zee Bridge; specific sites are described in **Table A1** of **Appendix A**.

### *2.2.2 Sample collection*

The sample locations were selected to target STP outfalls and tributaries that are expected to be major sources of micropollutants in the HRE. Grab samples were collected in collaboration with Riverkeeper,[78] an organization dedicated to monitoring and protecting the waters of the HRE, during nine sampling events over the 2016 and 2017 recreational seasons (see **Table A2** for a complete list of the sampling dates and times). Samples were collected in 1L amber, trace clean glass bottles and stored in an ice bath on the sampling vessel for up to three days. The samples were then shipped on ice in a cooler to our laboratory at the end of each sampling campaign, and stored at 4°C until sample preparation. The total sample holding time prior to sample preparation and analysis was always between one and four days. A full description of the sampling procedure is provided in **Appendix A**. The sample sites included three STP outfalls (Orangetown, O_STP; West Point, W_STP; and Rondout Creek, R_STP), four sites at the mouth of tributaries (Pocantico River, PR_M; Cedar Pond Brook, CB_M; Furnace Brook, FB_M; and Annesville Creek, AC_M), eight sites inside tributaries (Rondout Creek: RC_U (upstream), RC_D (downstream); Esopus Creek: EC_U, EC_D; Catskill Creek: CC_U, CC_D; Normans Kill, NK; and the Mohawk River, MR), and two control sites that were sampled in the mid-channel of the

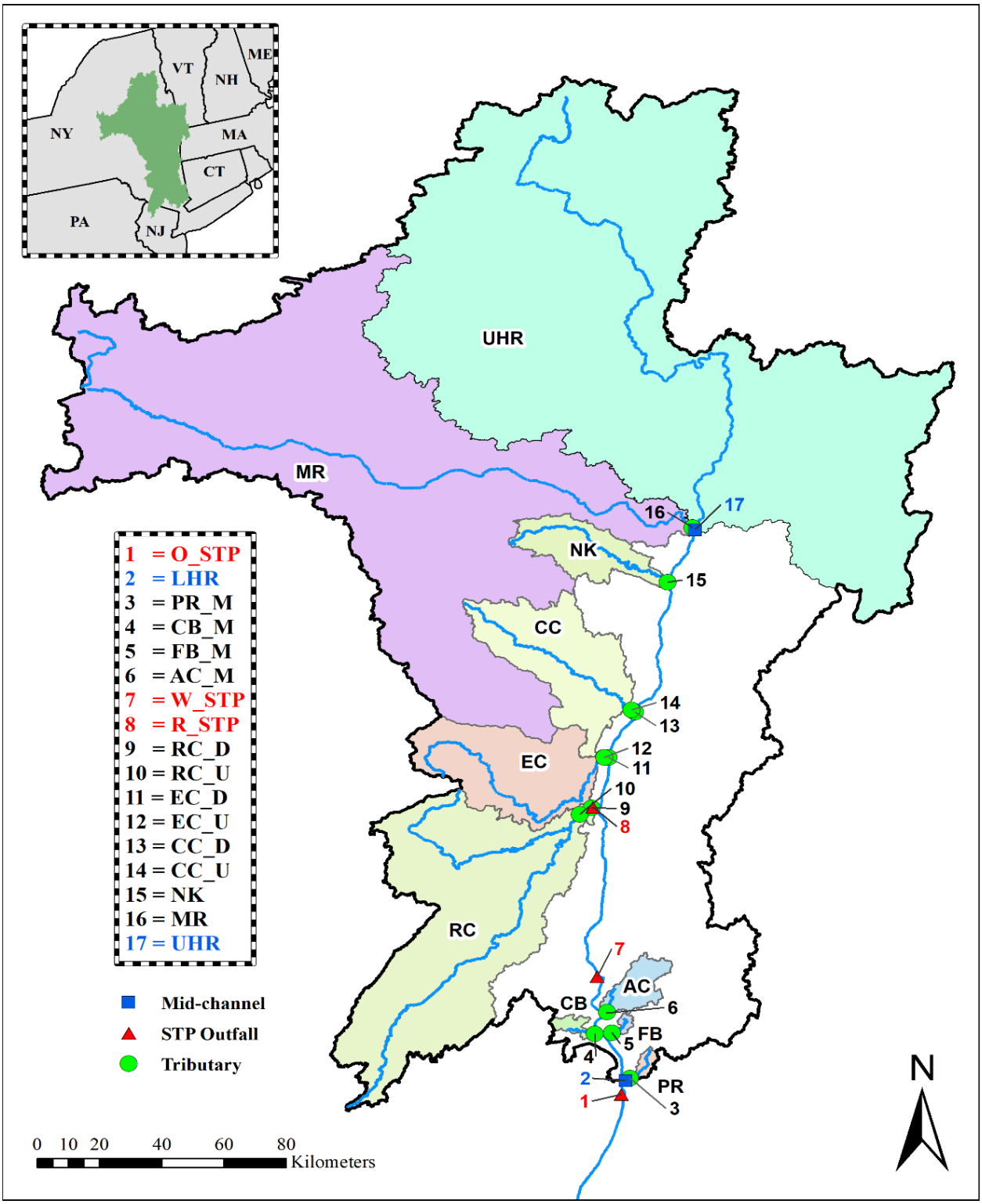



**Figure 2.1:** Map of the Hudson River Estuary catchment area and select tributary watersheds. The 17 sampling locations are marked with site IDs. The full name and description of each site is provided in **Table A1**.

HRE at the northern (Upper Hudson River, UHR) and southern (Lower Hudson River, LHR) ends of the study boundaries. The STP outfall samples were collected from the upwelling or directly adjacent to STP outfall pipes and thus contain a mixture of STP effluent and river water. Two samples were lost during sample shipment, therefore a total of 127 samples were processed and analyzed. It must be noted that data derived from grab samples do not necessarily reflect the expected dynamics of micropollutant occurrence or concentration in surface water systems.[38] However, a series of grab samples can be analyzed to provide robust estimates of the likelihood of occurrence and average concentrations of specific micropollutants at a particular sample site. We further note that no field blanks were collected during this study, though the sampling procedure was explicitly designed to limit contamination in the field. We converted measured micropollutant concentrations to loads using river flow data obtained from USGS stream gages under the assumption that the water columns were well-mixed.[79] The lower portion of the HRE is a partially-mixed estuary with significant vertical stratification.[80] Therefore, we only estimated loads from samples collected inside of tributaries that are located in the upper portion of the HRE where vertical stratification is not expected. USGS streamflow rates are provided in **Table A3**.

### *2.2.3 Sample preparation and analysis*

The samples were prepared using a mixed-bed solid-phase extraction (SPE) method to concentrate the samples as previously described.[81] We then used high performance liquid chromatography (HPLC) coupled to high-resolution mass spectrometry (MS) to quantify the occurrence of 200 diverse micropollutants which have been previously detected or are likely to occur in surface waters.[8,81] The analytical HPLC-MS/MS method was previously developed and validated for a broad range of micropollutants.[81,82] These methods are summarized in **Appendix A**

and the target micropollutants, their respective use-class, structure, physiochemical properties, analytical data, and limits of quantification (LOQs) are provided in **Tables A4 – A5**.

### *2.2.4 Geospatial analysis*

Mapping and geospatial analyses were conducted in ArcGIS v10.4. We used publically available data to produce maps of the HRE catchment area for geospatial references including·Land cover (**Figure A1**) and STP outfalls (**Figure A2**). The geospatial data sources are summarized in **Table A7**.

### *2.2.5 Statistical analysis*

Statistical analyses were conducted using R Statistical Software v3.3. Micropollutant clusters were determined using hierarchical clustering with the *hclust* function, Ward's agglomeration method, and either binary or Euclidean distance matrixes. Micropollutant data were converted into binary occurrence data (< LOQ = 0; ≥ LOQ = 1), and into normalized concentration data by z-score normalization to remove the effects of varying concentration differences at different sample sites.[36] Wilcoxon rank sum (WRS) tests, Tukey's honest significant difference (Tukey-HSD) tests, and analysis of variance (ANOVA) were used to assess the differences between measured micropollutant concentrations and estimated loads. We inspected Pearson correlations between individual micropollutant profiles using the *cor.test* function.

## 2.2 Results and Discussion

### *2.3.1 Micropollutant detection frequencies and concentrations*

We analyzed 127 samples collected from 17 sites at nine time points from May 2016 through October 2017. A summary of the number of micropollutants detected throughout the sampling campaign is provided in **Figure A3**. Micropollutants were defined by their primary use as either wastewater-derived or agricultural-derived, as described in **Table A4**; however, these are

not strict classifications. From our target list of 200 micropollutants, 168 were detected in at least one of the 127 samples; 116 were wastewater-derived and 52 were agricultural-derived micropollutants. Four of the 200 target micropollutants were detected in all 127 samples, and twenty micropollutants were detected in at least 120 samples. The micropollutants detected in all samples were atrazine (herbicide), gabapentin (antiepileptic), metolachlor (herbicide), and sucralose (artificial sweetener).

As expected, the highest concentrations of micropollutants were measured in the STP outfall samples, which contained a variable mixture of STP effluent and river water, depending on river flow and mixing conditions at the STP discharge. Sucralose, atenolol acid (metabolite of atenolol and metoprolol), and metformin (antidiabetic) were measured as high as the mid $mg \cdot L^{-1}$ range in the STP outfall samples, though dilution and mixing resulted in lower concentrations for these three micropollutants at adjacent downstream sampling sites ($ng \cdot L^{-1}$ to low $\mu g \cdot L^{-1}$ range). A majority of the remaining micropollutants (84%) were measured in the 1 – 100 $ng \cdot L^{-1}$ range, which is typical of concentrations measured in surface water systems around the world.[68] A full list of the micropollutant detection frequencies, and their measured minimum, median, and maximum concentrations is provided in **Table A9**.

### *2.3.2 Clustering based on spatiotemporal occurrence patterns*

We converted our micropollutant concentration data into binary occurrence data and used hierarchical clustering to cluster the micropollutants based on their spatiotemporal occurrence among all samples. The resulting dendrogram is presented in **Figure 2.2** and reveals three major clusters and four sub-clusters that define groups of micropollutants that have similar spatiotemporal occurrence patterns in the HRE. Previous studies have used principal component analysis or hierarchical clustering to identify groups of micropollutants that co-occur


**Figure 2.2:** Dendrogram of micropollutants clustered by spatiotemporal occurrence profiles in all samples. The dendrogram was cut (dashed line) to present three main clusters: core micropollutants (top, red), STP micropollutants (middle, blue), and diffuse micropollutants (bottom, green). These clusters were subdivided (dashed line) to present five total subclusters.

in the environment, but the breadth of micropollutants measured[35,36,39] or the spatiotemporal variability[32,83] among the samples in previous studies has been limited. Our data provide occurrence patterns for a relatively large number of micropollutants of varying use-classes measured with broad spatial and temporal resolution within a single watershed, providing a uniquely large dataset to examine micropollutant clustering. A summary of the cluster and sub-cluster to which each micropollutant was assigned is provided in **Table A9.**

We next sought to examine whether the major clusters or sub-clusters of micropollutants could be defined based on characteristic spatiotemporal occurrence patterns, geospatial references, and/or use-classes. The first major cluster (red cluster in **Figure 2.2**) contains 60 *core micropollutants* that were measured ubiquitously throughout the HRE in all sample types and at nearly all sample times. The *core micropollutants* cluster contains wastewater-derived micropollutants and agricultural-derived micropollutants, including the four micropollutants that were detected in all 127 samples. The mixed use-classes and ubiquitous occurrence of the *core micropollutants* suggest mixed but continuous sources within the HRE watershed and some level of environmental persistence among these micropollutants. The *core micropollutants* will be examined further by investigating normalized concentration patterns in the following section.

The second major cluster (blue cluster in **Figure 2.2**) contains 49 micropollutants that were frequently detected in STP outfall samples and sporadically in tributary or mid-channel control samples. The majority of micropollutants in this cluster (80%) are wastewater-derived micropollutants such as ranitidine (acid inhibitor) and sitagliptin (antidiabetic), but also include several residential-use pesticides like propoxur (insecticide) and mecoprop (herbicide) that are generally classified as agricultural-derived micropollutants. Based on the spatiotemporal occurrence pattern and use-classes of micropollutants in this cluster, we define this cluster as *STP*

*micropollutants*. Our geospatial analysis further reveals that *STP micropollutants* are most frequently measured in tributary samples that are located with less than 5 km of hydraulic distance from major STP outfalls (CB_M, RC_U, RC_D, EC_D, and NK); the other tributaries either have no major STPs within their watershed or have sampling sites that are more than 5 km downstream from the nearest major STP outfall. These data clearly highlight STP outfalls as important sources of *STP micropollutants,* but proximity to STP outfalls is also an important factor in the clustering of *STP micropollutants*.

The *STP micropollutants* cluster partitioned further into two distinct sub-clusters. The first sub-cluster contains micropollutants that were measured almost exclusively (> 91%) in STP outfall samples and was defined as the *STP exclusive* sub-cluster of micropollutants. This sub-cluster is also defined as containing micropollutants with the highest concentrations relative to the other spatiotemporal occurrence clusters and sub-clusters ($p < 0.05$, Tukey-HSD) and containing micropollutants with relatively high LOQs ($19 \pm 24$ ng·L$^{-1}$, avg. ± std.). Therefore, it is possible that *STP exclusive* micropollutants are present in some tributary or mid-channel control samples but are diluted to concentrations below the LOQ. This dilution effect is best exemplified by samples from W_STP which was sampled in the upwelling of the West Point military academy STP outfall. During sampling periods when the upwelling was not clearly visible due to higher river flow, the occurrence of *STP exclusive* micropollutants decreased by an average of 87%. The second sub-cluster contains micropollutants that were measured in STP outfall samples and sporadically in tributary or mid-channel control samples and was defined as the *wastewater (WW) sources* sub-cluster of micropollutants. *WW-sources* was further defined as containing micropollutants with relatively low LOQs ($4 \pm 5$ ng·L$^{-1}$) which may explain their detection in the diluted tributary and mid-channel control samples. Micropollutants in the *WW-sources* sub-cluster

may also be more persistent than micropollutants in the *STP exclusive* sub-cluster, resulting in their detection at locations distant from their source. We conclude that the presence of micropollutants in the *STP exclusive* sub-cluster is indicative of direct impact from STP outfalls, whereas the presence of micropollutants in the *WW-sources* sub-cluster may indicate nearby sewage influence (< 5 km) from STP outfalls, septic systems, or combined sewer overflows.

The third major cluster (green cluster in **Figure 2.2**) contains 56 micropollutants that were infrequently detected throughout the HRE in all sample types and were not necessarily influenced by STP outfalls. The majority of micropollutants in this cluster (68%) are agricultural-derived pesticides such as isoproturon (herbicide) and veterinary medicines such as sulfathiazole (antibiotic). Micropollutants in this cluster were measured at similar concentrations in the STP outfalls when compared to nearby tributary samples or mid-channel control sites ($p > 0.05$, paired WRS), suggesting that the sources of these micropollutants cannot be attributed to STPs and are likely derived from diffuse upstream sources. Based on these observations, we define this cluster as *diffuse micropollutants*.

Closer examination of this cluster allows us to further partition the micropollutants into two sub-clusters. The *Diffuse-specific* sub-cluster consists of 17 micropollutants that were only detected in specific STP outfall samples (O_STP in May 2016, and O_STP and W_STP in September 2016) and tributary samples (CB_M in May 2016 and EC_U in July 2016) during specific sampling events; due to the low occurrence of these micropollutants, we could not determine a likely source or mobilization process for this sub-cluster of micropollutants. The *Diffuse-general* sub-cluster consists of 39 micropollutants that were infrequently detected in all sample types and all sample times. Due to their limited occurrence and separate clustering from the *STP micropollutants* we can conclude that the sources of *Diffuse-general* micropollutants is

not limited to STP outfalls, but are also likely derived from diffuse upstream sources. The detection of *diffuse micropollutants* was rare and highlights the advantages of higher resolution temporal sampling which can capture more robust estimates of the occurrence of specific micropollutants at a particular sample site.

### *2.3.3 Clustering based on normalized concentration pattern.*

Due to the high detection frequency of *core micropollutants* throughout the HRE, this cluster was subjected to further investigation. We used z-score normalization to remove the effect of varying concentrations at different sample sites to investigate the spatiotemporal normalized concentration patterns among the *core micropollutants.*[36] The resulting dendrogram and heatmap is provided in **Figure 2.3** which reveals two distinct clusters of sample sites (horizontal axis) and four distinct sub-clusters of the *core micropollutants* (vertical axis). The sample sites clearly separated between tributary samples and STP outfall samples, which highlights the differences in the concentrations of *core micropollutants* between those sample types. Moreover, several pairs of parent compounds and transformation products were closely clustered, which indicates their co-occurrence at similar normalized concentrations across the sample sites. The four *core micropollutant* sub-clusters represent different spatiotemporal normalized concentration patterns. A summary of the sub-cluster to which each of the 60 *core micropollutants* is assigned is provided in **Table A9**.

The first sub-cluster (*sub-cluster A* in **Figure 2.3**) contains seven *core micropollutants* including atrazine and its transformation products, metolachlor, and metalaxyl. The concentrations of *sub-cluster A* micropollutants were not significantly different in tributary samples and STP outfall samples ($p > 0.05$, paired WRS). This suggests that the sources of *sub-cluster A* micropollutants, like the *diffuse micropollutants* cluster, cannot be exclusively attributed to STP

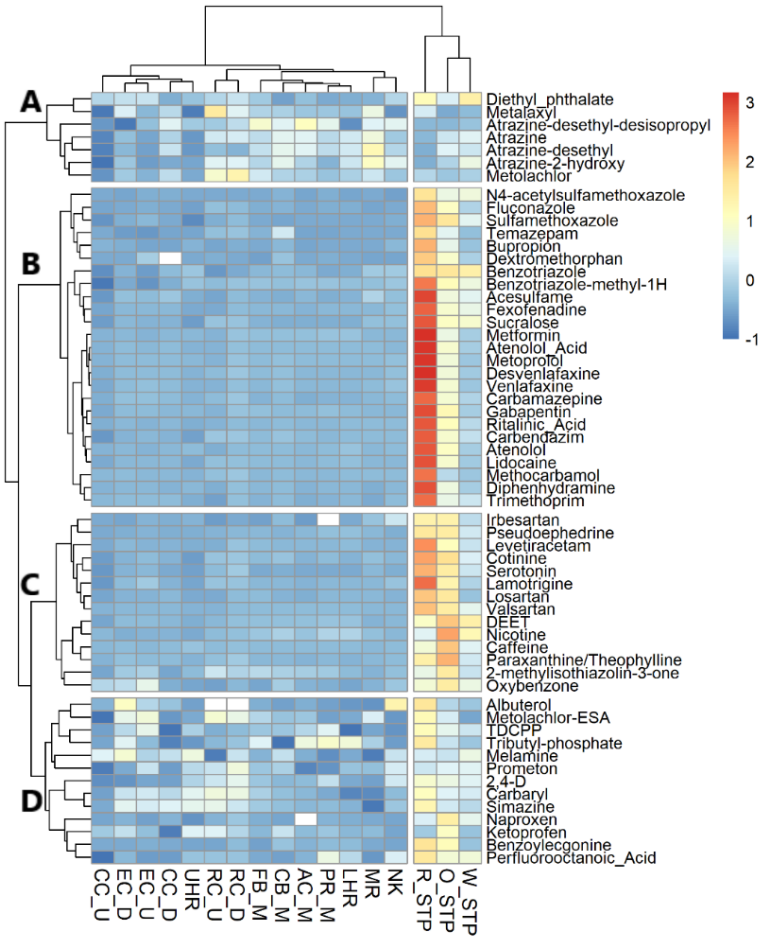



**Figure 2.3:** Heatmap of core micropollutants clustered by spatiotemporal z-score normalized concentration profiles in all samples. The color of each cell represents the average normalized concentration value. The dendrogram was cut to present two sample site clusters (tributaries and STP outfalls) and four micropollutant subclusters (A–D).

outfalls. We conclude that the micropollutants in *sub-cluster A* are predominantly used in agriculture or on urban landscapes and are attributed to diffuse upstream sources.[35,40]

The normalized concentration patterns of *sub-cluster A* micropollutants also separated the tributary sample sites into two main groups. The first group of tributary sample sites consists of UHR, EC, and CC and the second group consists of RC, MR, NK, and the five samples sites located in the lower HRE (LHR, PR_M, CB_M, FB_M, and AC_M). We examined the major differences in land cover in the watersheds of these groups of tributaries. Our geospatial analysis revealed that the watersheds of the former group has a significantly ($p < 0.01$, student's t-test)

lower extent of land cover defined as hay/pasture and cultivated cropland than the RC, MR, and NK watersheds. The increased extent of hay/pasture and cultivated cropland in the latter watersheds likely leads to the higher normalized concentrations of *sub-cluster A* micropollutants in these tributary watersheds. Furthermore, MR was the only tributary with significantly higher concentrations of *sub-cluster A* micropollutants when compared to UHR ($p < 0.01$, paired WRS); approximately 25% of the land cover in the MR watershed is either hay/pasture or cultivated cropland, which represents the highest agricultural land cover of any tributary watershed.

The next sub-clusters (*sub-cluster B* and *sub-cluster C* in **Figure 2.3)** contain 39 *core micropollutants* which occurred at significantly higher concentrations in the STP outfall samples when compared to the other sampling sites ($p < 0.01$, paired WRS). S*ub-cluster B* micropollutants were present at higher concentrations in the STP outfall samples than *sub-cluster C* micropollutants ($p < 0.01$, WRS), though there was not a significant difference between the concentrations among the tributary samples ($p > 0.05$, WRS). As expected, these sub-clusters contain almost entirely wastewater-derived micropollutants (37 of 39), but each sub-cluster also contains one micropollutant that is often identified as an agricultural-derived micropollutant. First, 2-methylisothiazolin-3-one (MI) is a biocide that is used in agriculture but is also an ingredient in personal care products.[84] Our results demonstrate that the main sources of MI to the HRE are STP outfalls, suggesting primary use in personal care products. Second, carbendazim (fungicide) is banned for use as an agricultural pesticide in the United States, but is still used in paints, building materials, and other products.[85] Carbendazim is also a transformation product of thiophanate-methyl, which is registered in the United States as a residential and crop fungicide.[86] In a recent study, carbendazim was found at higher concentrations at sites with higher urban land cover.[87] Here, carbendazim was detected in 92% of our samples and our higher resolution sampling allows

us to identify the major source as STP outfalls, rather than diffuse upstream sources or higher urban land cover.

The HRE was found to be more impacted by *sub-clusters B* and *C* micropollutants as it flows south. These micropollutants had significantly higher concentrations in LHR than UHR, despite dilution from the increased flowrate ($p < 0.01$, paired WRS). Our results indicate that only *sub-clusters B* and *C* micropollutants have concentrations that increase as the HRE flows south. Additionally, *sub-clusters B* and *C* micropollutants were detected at significantly higher concentrations in certain tributaries with major STP outfalls (RC_U, EC_U, NK, and MR) when compared to UHR and at all STP outfalls when compared to both control sites ($p<0.01$, paired WRS). Interestingly, CC_U was found to have significantly lower concentrations of *sub-clusters B* and *C* micropollutants when compared to the control sites ($p < 0.01$, paired WRS) and the CC watershed has no major STPs. Therefore, we conclude that s*ub-clusters B* and *C* micropollutants are accumulating in the HRE through STP outfalls or septic systems within tributary watersheds.

The last sub-cluster (*sub-cluster D* in **Figure 2.3**) contains thirteen *core micropollutants* representing a mixture of wastewater-derived and agricultural-derived micropollutants. *Sub-cluster D* micropollutants were found at significantly higher concentrations in the STP outfall samples than in tributary samples ($p < 0.01$, paired WRS), but at significantly lower concentrations than *sub-clusters B* and *C* micropollutants in STP outfall samples ($p < 0.01$, WRS). Concentrations of *sub-cluster D* micropollutants in tributary samples were significantly higher than *sub-cluster A* ($p < 0.01$, WRS), but similar to *sub-clusters B* and *C* ($p > 0.05$, WRS). The relatively high normalized concentrations of *sub-cluster D* micropollutants in the tributary samples suggest that they have mixed-sources and/or are more persistent in the environment than the other *core micropollutants*. Several agricultural-derived micropollutants in this cluster are used both

commercially and residentially (for example, 2,4-D, carbaryl, prometon, and simazine). Additionally, the three pharmaceuticals in this cluster are used as both human and veterinary medicines. Based on these observations, we conclude that *sub-cluster D* represents a group of micropollutants with mixed-uses and multiple sources.

### *2.3.4 Micropollutant loads and relative contribution of tributaries*

We next aimed to evaluate whether there were any temporal patterns to micropollutant loading in the HRE and to identify the relative contributions of each tributary to total micropollutant loadings in the HRE. We calculated cumulative micropollutant loads from samples taken inside of the tributaries (RC_U, EC_U, CC_U, NK, MR, and UHR) for which we had access to streamflow data.[35,88] We first examined the cumulative loads of each micropollutant cluster and sub-cluster in each month across each of the sampling campaigns. Among the three major clusters of micropollutants, only the loadings of the *diffuse micropollutants* were found to change significantly as a function of sample month ($p < 0.01$, ANOVA) which suggests a temporal dynamic to their input into the HRE. More specifically, micropollutants in the *Diffuse-general* sub-cluster were detected at significantly higher loads in samples collected from May through July. Even though *Diffuse-general* micropollutants were detected in only about half of the total samples, the cumulative loads of *Diffuse-general* micropollutants in the HRE during our sampling events were comparable to the *core micropollutants*. In fact, the cumulative loads of the *Diffuse-general* micropollutants were significantly correlated with flowrate ($p < 0.01$, $\rho = 0.76$, Pearson correlation) which suggests transport in runoff from agricultural or urban landscapes during a limited application period. There were also notable temporal dynamics among *sub-cluster A* of the *core micropollutants*. Like the *Diffuse-general* micropollutants, loadings of *sub-cluster A* micropollutants are also elevated in samples collected from May through July. However, unlike

the *Diffuse-general* micropollutants, *sub-cluster A* micropollutants were also detected throughout the entire recreational season, suggesting increased persistence of *sub-cluster A* micropollutants relative to the *Diffuse-general* micropollutants. Two *sub-cluster A* micropollutants, atrazine and metolachlor, have been frequently detected in surface waters beyond typical application seasons, providing some validation to these observations.[39,89]

We next sought to identify the relative contributions of each tributary to total micropollutant loadings in the HRE. Because we assumed that the water columns were well-mixed, the high average flowrates of MR (101.3 $m^3 \cdot s^{-1}$) and UHR (179.0 $m^3 \cdot s^{-1}$) compared with the other tributaries (2.0 – 7.6 $m^3 \cdot s^{-1}$) translate into significantly higher estimated loadings of all micropollutants in these sub-watersheds. Nevertheless, we can conclude that MR contributes greater cumulative loadings of *STP micropollutants* and *sub-clusters A, B,* and *C* micropollutants among the *core micropollutants* than UHR. Conversely, UHR contributes greater cumulative loadings of *diffuse micropollutants* and *sub-cluster D* micropollutants among the *core micropollutants* to the HRE than MR. Our geospatial analysis of these two sub-watersheds support these observations. The MR watershed has greater extents of agricultural land cover and more major and minor STP outfalls than the UHR, which can explain the greater loadings of the respective clusters and sub-clusters of micropollutants associated with the MR. The greater flowrate of the UHR explains the greater loadings of *diffuse micropollutants* and *sub-cluster D* micropollutants among the *core micropollutants* which are present in both sub-watersheds at similar concentrations ($p > 0.05$, paired WRS).

The remaining tributaries for which we had access to streamflow data (RC, EC, CC, and NK) had lower but similar flowrates, which allowed us to examine relative contributions of micropollutant loadings among them. We conclude that RC contributes greater cumulative

loadings of *STP micropollutants* and *sub-clusters A, B,* and *C* micropollutants among the *core micropollutants* than the other tributaries. RC also has significantly greater extents of agricultural land cover and more major and minor STP outfalls than the other sub-watersheds which support this observation. Despite having similar average streamflow rates as RC (6.8 $m^3 \cdot s^{-1}$), CC (7.6 $m^3 \cdot s^{-1}$) and EC (5.1 $m^3 \cdot s^{-1}$) contribute insignificant cumulative micropollutant loadings to the HRE. Finally, even though NK was found to have higher concentrations of *STP micropollutants* and *sub-clusters B* and *C* micropollutants than any of the other tributaries, the low average streamflow rate (2.0 $m^3 \cdot s^{-1}$) resulted in insignificant micropollutant loadings from NK to the HRE.

### *2.3.5 Predictors of micropollutant occurrence*

Our relatively large dataset allows for additional statistical analyses that may inform future water quality monitoring studies. We first explored relationships among summary statistics including the total number of detections of micropollutants, the cumulative concentration of micropollutants, and the concentrations of individual micropollutants measured in each sample. We focused our investigation on cumulative concentrations instead of loads so that we could fully utilize our dataset and not be restricted to samples for which streamflow was available. As expected, we observed a significant relationship between the total number of detections and the cumulative concentration of micropollutants in each sample ($p < 0.01$, $\rho = 0.72$, Pearson correlation).[32] However, the data regression shown in **Figure 2.4** reveals four groups of samples that significantly deviate from the overall trend. The first group of two samples can be categorized as having a low number of detections and a low cumulative concentration. Both samples are from CC_U and were among the cleanest samples measured in the sampling campaign in terms of both concentration and cumulative loads. The low concentrations measured in these samples could be due to the low agricultural impact of the watershed and the lack of upstream major STP outfalls.

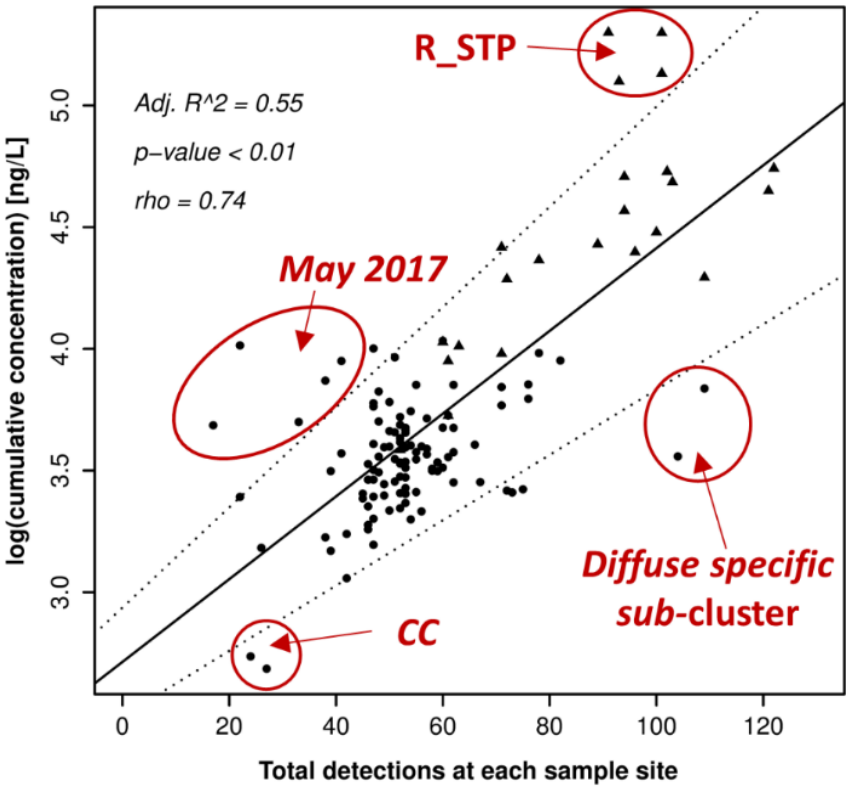



**Figure 2.4:** Linear regression (solid line) of cumulative concentration and total detections of micropollutants at each sample site (●, tributaries; ▲, STP outfalls). Dashed lines represent the 99% confidence interval.

The next group of two samples can be categorized as having a high number of detections and a low cumulative concentration. These two samples are from CB_M in May 2016 and EC_U in July 2016, which are the same samples that determined the *Diffuse-specific* sub-cluster of micropollutants. These samples contained 17 rare micropollutants that were present at relatively low concentrations, explaining the deviation from the overall trend. The next group of four samples can be categorized as having a high number of detections and a high cumulative concentration. These four samples are from R_STP the high concentrations of metformin and sucralose in these samples ($mg \cdot L^{-1}$ range) control the deviation from the overall trend. Finally, and perhaps most interestingly, the remaining group of five samples can be categorized as having a low number of detections and a relatively high cumulative concentration. These five samples are all from the May

2017 sampling event and represent five northern sampling sites in the HRE (UHR, NK, CC_D, EC_U, and EC_D). The deviation from the overall trend is the result of elevated concentrations of perfluorobutanoic acid (PFBA), an industrial perfluorochemical. Emerging concerns over PFASs in water resources has led to the recent development of shorter-chain analogues[90] and there is limited knowledge regarding their major sources to the environment.[36] PFBA is also a transformation product of other more complex PFASs.[91] Our geospatial analysis revealed the possible source of the spike as a major international airport—concentrations of PFBA are elevated downstream of the airport and decrease as the HRE flows south, and lower concentrations were found in samples taken upstream inside of the tributaries. These observations demonstrate how an analysis of trends and outliers in summary statistics can be used to identify a potential contamination event within large water quality datasets.

We also observed a significant relationship between the total number of detections or the cumulative concentration of micropollutants in each sample and the concentrations of some individual micropollutants ($n > 50$, $p < 0.01$, $\rho > 0.65$, Pearson correlation).[32] **Table A10** identifies 17 micropollutants whose concentrations are significantly associated with the total number of micropollutant detections (top three are dextromethorphan, naproxen, and carbendazim) and 13 micropollutants whose concentrations are significantly associated with the cumulative concentrations of all detected micropollutants (top three are sitagliptin, bupropion, and naproxen). Interestingly, the individual micropollutants that are the best predictors of overall micropollutant occurrence and concentrations are mostly *core micropollutants;* specifically, 70% of the best predictors were classified in *sub-clusters B* or *C.* In a recent micropollutant survey conducted across the United States, the USGS identified desvenlafaxine, cotinine, lidocaine, and metformin as good predictors of overall micropollutant occurrence.[32] Each of these micropollutants were also

identified as significant predictors in our study, providing some evidence that our results and conclusions can be extrapolated to other surface water systems. A number of other studies have proposed micropollutant indicators of wastewater or other anthropogenic influence,[92] though none have used a clustering approach to link micropollutant indicators to groups of micropollutants with similar sources, transport behavior, and persistence behavior at the watershed scale.

Finally, we detected eight pairs of parent compounds and transformation products (TPs) classified as *core micropollutants*. Despite varying physiochemical properties, the concentrations of six of these pairs (venlafaxine – desvenlafaxine; benzotriazole-methyl-1H – benzotriazole; caffeine – paraxanthine/theophylline; atrazine – atrazine-2-hydroxy and atrazine-desethyl; atenolol and metoprolol – atenolol-acid; and sulfamethoxazole – N4-acetyl sulfamethoxazole) were significantly and linearly correlated (p-value $< 0.01$, $R^2 > 0.60$, $\rho > 0.80$, Pearson correlation). Linear regressions between the parent compounds and TPs are provided in **Figure A4**. These pairs were also closely clustered based on normalized concentrations, as shown in the spatiotemporal z-score normalized concentration heatmap (**Figure 2.3**). These data suggest that the occurrence and concentration of these parent compounds are good predictors of the occurrence and concentration of the associated TP. However, the concentrations of two other pairs of parent compounds and TPs (nicotine – cotinine; and metolachlor – metolachlor-ESA) showed significant but poor correlation ($p < 0.01$, $R^2 < 0.30$, $\rho < 0.6$, Pearson correlation). The poor correlation for these compounds could be explained by either additional major TPs that were not included in our study or varying sources of the parent micropollutant and TP pairs. For example, metolachlor is known to be transformed to both metolachlor-ESA and metolachlor-OXA in variable proportions and we only included metolachlor-ESA in our study.[93] Further, metolachlor is typically sourced to runoff from

agricultural landscapes whereas metolachlor-ESA may be sourced to either agricultural runoff or groundwater recharge.[94]

### *2.3.6 Environmental Implications*

The data reported in this study represent the first extensive investigation of micropollutants in the HRE. These data also provide a uniquely large dataset with which we were able to perform robust geospatial analyses to identify clusters of micropollutants that are present with similar spatiotemporal occurrence patterns or normalized concentration patterns throughout the HRE. These clusters were linked to a variety of sources based on the types of micropollutants contained in the clusters and the geospatial features of the watersheds in which they were measured.

We contend that the compositions of the micropollutant clusters that we report are conserved across surface water systems throughout the world. To examine this, we reviewed the literature on micropollutant monitoring of surface water systems reported around the world. We found that the most frequently detected micropollutants in other studies, most notably a national monitoring campaign of US streams[32] and an EU-wide survey,[95] are identified as *core micropollutants* in this study (atrazine, benzotriazole, benzotriazole-methyl-1H, caffeine, carbamazepine, cotinine, DEET, desvenlafaxine, lidocaine, metformin, metolachlor, metolachlor-ESA, PFOA, and sulfamethoxazole). As another example, a recent study in Japan reported *sub-cluster B* and *C* micropollutants at higher frequencies near STP outfalls (atenolol, carbamazepine, metoprolol, theophylline, and trimethoprim) and *STP micropollutants* were more frequently detected in the STP outfalls than nearby river samples (ciprofloxacin, diclofenac, furosemide, norfloxacin, ofloxacin, and propranolol).[96] The same study also reported no or variable detections of *diffuse micropollutants* (clofibric acid, naproxen, and sulfathiazole) not related to STP outfalls.[96] A study conducted in a small US watershed revealed nearly ubiquitous occurrence of several *core*

*micropollutants* (caffeine, cotinine, DEET, and trimethoprim) and linked several *diffuse micropollutants* and *sub-cluster A* micropollutants (acetochlor, atrazine, and metolachlor) to diffuse upstream sources.[35] As a final example, a study examining micropollutant occurrence in receiving waters downstream from STP outfalls revealed nearly ubiquitous occurrence of several *core micropollutants* (atenolol, caffeine, carbamazepine, cotinine, DEET, diphenhydramine, paraxanthine, sulfamethoxazole, trimethoprim, and valsartan) and decreasing concentrations of *STP micropollutants* (gemfibrozil and ibuprofen) at samples collected further downstream from the STP outfall.[97] Together, these data and others collected from similar studies support the idea that the micropollutant clusters discovered in this study can be used to characterize micropollutant occurrence and sources in surface water systems around the world.

Analysis of our dataset also revealed a number of micropollutants whose concentrations can be used to predict overall micropollutant occurrence or cumulative micropollutant concentrations in the HRE. Although these relationships may need to be recalibrated to be applied quantitatively across surface water systems, we expect that relative measurements of indicator micropollutants identified in this study can be used to assess the relative extent of micropollutant occurrence and concentrations in other surface water systems.

Finally, we identified specific tributaries that contribute varying loads of micropollutants to the HRE. Although the UHR and MR are obvious contributors of micropollutant loads to the HRE due to their size, we found that RC contributed significantly greater loads of micropollutants to the HRE than the other smaller tributaries. Water quality stakeholders interested in minimizing the concentrations of micropollutants in the HRE should focus on these three sub-watersheds, along with STP outfalls, for implementation of mitigation strategies.

# CHAPTER 3 – Fall Creek Monitoring Station: Highly Resolved Temporal Sampling to Prioritize the Identification of Nontarget Micropollutants in a Small Stream[b]


## Abstract

The goal of this research was to comprehensively characterize the occurrence and temporal dynamics of target and nontarget micropollutants in a small stream. We established the Fall Creek Monitoring Station in March 2017 and collected daily composite samples for one year. We measured water samples by means of high-resolution mass spectrometry and developed and optimized a post-acquisition data processing workflow to screen for 162 target micropollutants and group all mass spectral (MS) features into temporal profiles. We used hierarchical clustering analysis to prioritize nontarget MS features based their similarity to target micropollutant profiles and developed a high-throughput pipeline to elucidate the structures of prioritized nontarget MS features. Our analyses resulted in the identification of 31 target micropollutants and 59 nontarget micropollutants with varying levels of confidence. Temporal profiles of the 90 identified micropollutants revealed unexpected concentration-discharge relationships that depended on the source of the micropollutant and hydrological features of the watershed. Several of the nontarget micropollutants have not been previously reported including pharmaceutical metabolites, rubber vulcanization accelerators, plasticizers, and flame retardants. Our data provide novel insights on the temporal dynamics of micropollutant occurrence in small streams. Further, our approach to nontarget analysis is general, and not restricted to highly resolved temporal data acquisitions or samples collected from surface water systems.



[b] Reproduced with permission from *Environmental Science & Technology,* 2019, *53* (1), 77-78. DOI: 10.1021/acs.est.8b05320. Copyright 2019 American Chemical Society.

## 3.1 Introduction

Organic micropollutants are anthropogenic organic chemicals that are present in the environment at low concentrations.[1] Decades of monitoring studies have identified hundreds of organic micropollutants in natural and engineered water systems.[3,98] However, conventional micropollutant monitoring strategies focus on a finite number of target micropollutants,[32] a practice which is known to underestimate micropollutant exposure risk by a factor of 2 – 10, even when considering just one class of micropollutant (*e.g.*, pesticides).[33] Additionally, infrequent grab samples do not adequately capture the expected temporal dynamics of micropollutant concentrations (*e.g.*, seasonal runoff from agriculture) and will likely miss peak events associated with the greatest ecological risks. This problem is exacerbated in small streams, which are more sensitive to changing hydrologic conditions and have lower dilution rates than larger rivers.[99–101] Additionally, more than 110 million people in the U.S. are supplied by public drinking water systems that rely at least in part on small streams (defined as intermittent, ephemeral, and headwater streams),[102] though micropollutant monitoring in small streams has been rather limited.[40,103] New approaches are needed to more comprehensively characterize the occurrence and temporal dynamics of micropollutants in water systems, and particularly in small streams.

One way to broaden the scope of micropollutant monitoring is to complement conventional targeted screening with nontargeted screening techniques. Nontargeted screening by means of high-resolution mass spectrometry and post-acquisition data processing has emerged as an effective tool to comprehensively characterize the occurrence of organic chemicals in a variety of sample types including sediments,[41] animal fat,[54] dust,[55] wastewater effluent,[104] and surface water.[34] However, two major challenges still limit the widespread use of nontargeted screening for routine micropollutant monitoring. First, high-resolution mass spectral acquisitions contain

thousands of nontarget MS features. Prioritization of the most relevant nontarget MS features for structural elucidation is essential.[52] Most previously reported prioritization strategies are driven by analytical or chemical data; for example, others have prioritized nontarget MS features based on relative peak intensity,[56] the presence of strong isotope signatures (*e.g.*, chlorine atoms),[105] or evidence that a nontarget MS feature is part of a homologous series.[60] Whereas these prioritization strategies have led to the successful identification of nontarget micropollutants, prioritization strategies that are coupled with features of the system being studied may lead to more generalizable conclusions about micropollutant occurrence and temporal dynamics. Second, there is no broadly accepted approach to elucidate structures of nontarget MS features. A variety of vendor software has emerged in recent years, though proprietary algorithms for structural elucidation do not afford the transparency needed to fully evaluate annotations of nontarget MS features. A number of open source tools[53,64,106–108] and databases[109–112] have also been developed to address certain aspects of nontarget analysis and structural elucidation, though there is a need to develop and optimize data analysis pipelines to enable high throughput structural elucidation of nontarget MS features.

Another way to gain insights on micropollutant occurrence and temporal dynamics in a water system is to implement a more continuous sampling strategy. For example, intermittent sampling of surface water systems around the U.S. over several years revealed distinct temporal profiles of pesticides that peaked during the agricultural growing season.[113] Daily composite sampling of a surface water collection system revealed that antecedent and post-application rain events trigger glyphosate transport from runoff-prone soils.[114] Recent studies in the Rhine River have demonstrated that daily composite sampling over long periods of time can reveal unexpected temporal dynamics of target micropollutants[37] and trends in nontarget MS features can be used to detect contamination events.[57]

The objectives of this study were to: (i) establish a continuous monitoring station in a small stream to generate highly resolved temporal profiles of target micropollutants and nontarget MS features; (ii) use the highly resolved temporal profiles to prioritize nontarget MS features for structural elucidation; and (iii) explore the highly resolved temporal data to reveal the temporal dynamics of micropollutant occurrence and gain fundamental insights on contaminant sources, fate, and transport phenomena. We selected the drinking water intake on Fall Creek (Ithaca, NY) as the location of the monitoring station. We collected daily composite samples from Fall Creek for one year and measured the samples by means of high-resolution mass spectrometry. The data were used to develop and optimize a nontarget MS feature prioritization workflow and structural elucidation pipeline using open source tools. Our approach led to the discovery of several types of micropollutant temporal profiles, some of which exhibited strong concentration-discharge dependencies. We detected 31 target micropollutants and elucidated the structures of 59 nontarget micropollutants with varying levels of confidence. These data provide novel insights on the temporal dynamics of micropollutant occurrence in small streams and the most comprehensive assessment of polar organic micropollutant exposure that is presently possible.

## 3.2 Material and Methods

### *3.2.1 Study area*

We selected the drinking water intake located on Fall Creek as the location of the Fall Creek Monitoring Station (FCMS). Fall Creek is a small tributary of Cayuga Lake located in Ithaca, New York and is the source of drinking water for over 30,000 people. The Fall Creek watershed upstream of the FCMS has an approximate area of 320 $km^2$ and over 22,000 people live within the watershed boundaries. A U.S. Geological Survey (USGS) stream gage located less than

2 km downstream from the FCMS recorded an average discharge of 7 $m^3 \cdot s^{-1}$ (median 4.0 $m^3 \cdot s^{-1}$, range 0.7 – 70 $m^3 \cdot s^{-1}$) during the study period,[79] reflecting both the small size of Fall Creek and variable streamflow. A GIS analysis of the watershed revealed that 29% and 16% of the land cover is defined as pasture/hay and cultivated crops, respectively.[115] Additionally, two sewage treatment plants (STPs) discharge directly into Fall Creek and 45% of the population utilizes onsite wastewater treatment and disposal systems.[116,117] Therefore, Fall Creek likely receives intermittent loadings of a variety of agricultural and STP-derived micropollutants. A map of the study area is provided in **Figure B1** of **Appendix B**.

### *3.2.2 Sample collection*

We used an ISCO automatic sampler (6712 Full-Size Portable Sampler, Teledyne Isco) to collect daily, time-proportional composite samples directly from the raw water intake of the Cornell Water Filtration Plant. We collected approximately 1 L daily samples through Teflon-lined polyethylene tubing in 1.8 L glass bottles using a 20 minute sampling interval. Additionally, we obtained weekly field blanks by collecting 1 L of nanopure water through the automatic sampler. Teflon-lined polyethylene tubing was replaced approximately every four months for precautionary purposes. We retrieved the samples from the FCMS at weekly intervals and stored them at 4°C until preparation, which was always within 24 hours of retrieval. Daily samples and weekly field blanks were collected between March 2017 – 2018.

### *3.3.3 Standards and reagents*

Details on the sources, preparation, and storage of authentic reference standards and reagents are provided elsewhere.[98] MS acquisition parameters for 162 target micropollutants and 33 isotope labeled internal standards (ILISs) are provided in **Tables B1** and **B2**.

### *3.3.4 Sample preparation and analysis*

We prepared the samples and field blank at weekly intervals by transferring 45 mL of each sample into a 50 mL conical tube (VWR), centrifuging at 4700 RPM (4816 g) for 15 minutes (Sorvall Legend XTR, Thermo Scientific), amending with 0.1% (v/v) 1 M ammonium acetate buffer, and adjusting the pH to 6.5 ± 0.2 using 5% formic acid and 1.4 N ammonia. Exactly 8 mL of each pH-adjusted sample was then transferred into triplicate 10 mL glass sample vials (Chromacol, Thermo Fisher Scientific). We also prepared one quality control sample each week by diluting a mixture of reference standards to 250 $ng \cdot L^{-1}$ in nanopure water. Each sample, field blank, and quality control sample was then spiked with 2 ng each of a mixture of 33 ILISs and stored at 4°C until analysis, typically within one week of preparation.

We adapted a previously described analytical method that implements large volume injection (LVI) and high performance liquid chromatography (HPLC) coupled with high-resolution mass spectrometry (QExactive hybrid quadrupole orbitrap, Thermo Fisher Scientific).[81,118] Briefly, samples were injected at 5 mL volumes onto a Hypersil GOLD aQ trap column (2.1 x 20 mm, particle size 12 µm, Thermo Fisher Scientific) at room temperature (21 – 22°C) and eluted with a mobile phase gradient onto an XBridge C-18 analytical column (2.1 x 50 mm, particle size 3.5 µm, Waters) at 25°C for analyte separation. Full scan mass spectra were acquired in positive ionization mode at a resolution of 140,000 at 200 *m/z*. Data dependent $MS^2$ spectra were acquired at the exact masses and retention times of all target micropollutants and prioritized MS features (see details in the following sections); the data dependent $MS^2$ inclusion list was continuously updated throughout the duration of the study. Additional data dependent $MS^2$ spectra were acquired for the most intense MS features if the inclusion list was not triggered. A total of three $MS^2$ scans were recorded after each full scan. Additional details on the analytical

method including the mobile phase gradient and the MS and $MS^2$ acquisition parameters are provided in **Tables B3** and **B4**. We note that the sample preparation and analytical methods selected for this work constrain the scope of our nontarget analysis to include polar to semi-polar organic molecules that can be ionized in positive mode electrospray ionization.

### *3.3.5 Peak picking and profile generation*

We developed and optimized an automated workflow for the characterization of target micropollutants and nontarget MS features using enviMass v3.413.[53] Our workflow consists of nine steps: (1) convert instrument .RAW files into .mzXML files using ProteoWizard v3.0.10827;[119] (2) identify fully-resolved chromatographic peaks using the peak picking settings provided in **Table B5** and assign a mass-to-charge ratio (*m/z*), a retention time (RT), and an intensity to each of the picked peaks; (3) recalibrate the *m/z* of each of the picked peaks based on the measured *m/z* of the ILISs in each sample (controls for mass drift); (4) exclude picked peaks that are not present in all three of the triplicate sample injections; (5) identify picked peaks in the field blank that have matching *m/z* and RT with picked peaks in each sample and calculate a sample-to-blind intensity ratio; (6) annotate any target micropollutants and ILISs based on expected *m/z*, RT, and isotopic signature; (7) group picked peaks with the same *m/z* and RT across samples into profiles; (8) normalize the intensities of all profiles based on the measured intensities of the ILISs (controls for variable matrix effects); and (9) identify isotopologues and adducts associated with a parent chemical and group into components. The output from this workflow is a final profile list of MS features described by their average *m/z* (tolerance set at 3 ppm) and RT (tolerance set at 30 s) among all of the samples, and their ILIS normalized intensity in each sample (a surrogate for concentration). Each MS feature in the profile list is further annotated as a target micropollutant, an ILIS, or a nontarget MS feature. We optimized the enviMass settings provided

in **Table B5** to maximize the number of true positive micropollutant annotations using data acquired from a quality control sample and the first several weeks of daily samples obtained from the FCMS. We first performed a conventional target screening using XCalibur v3.1 (Thermo Fisher Scientific) as previously described,[98] and then compared the results with the micropollutant annotations from our enviMass workflow. All settings were iteratively optimized to maximize the number of true positive micropollutant annotations in the quality control sample and to generate identical micropollutant profiles between data processed by XCalibur and enviMass among the samples from the FCMS.

### *3.3.6 Filtering and clustering of MS feature profiles*

As a means to prioritize profiles of nontarget MS features for structural elucidation, we first applied a series of data reduction filters including thresholds for sample-to-blind ratio ($\geq 10$), mean trend intensity ($\geq 10^5$), and RT ($\geq 6.5$ min) assigned to each MS feature profile. We also excluded all MS feature profiles associated with lower order isotopologues and adducts. Finally, we excluded all MS feature profiles that did not contain at least 30 total detections and at least 10 consecutive detections among the annual daily samples. We then grouped the remaining MS feature profiles by means of a hierarchical clustering analysis (HCA) using the *hclust* function in the R Statistical Software v3.3.3.[120] using Ward's agglomeration method and Euclidean distance matrices[41,98] based on the similarity of their ILIS normalized intensities over time; the ILIS normalized intensities of each MS feature profile were further normalized to their maxima so that the profiles were clustered based on their temporal trends and not absolute intensity. Nontarget MS features that clustered within or adjacent to localized clusters containing target micropollutants were prioritized for structure elucidation.

### *3.3.7 Structural elucidation of nontarget MS features*

We developed and optimized an automated pipeline to assign chemical structures to the prioritized nontarget MS features using a series of self-written R scripts and publically available R packages. We note that high-resolution mass spectrometry alone can lead to putative structural assignments by means of spectral annotation, but multiple analytical techniques are required to unequivocally annotate the structure of an unknown chemical. Therefore, all chemical structures are assigned a confidence level based on previously established criteria summarized in **Table B6**.[61] The accurate masses (*m/z*) assigned to each of the prioritized nontarget MS features are considered to be exact masses of interest (level 5). For each nontarget MS feature profile, the sample with the highest ILIS normalized intensity is selected and the package RMassBank[65] is used to extract MS and $MS^2$ data into R. The package GenFormR[121] is then used to predict molecular formulae based on MS isotopic signatures, $MS^2$ fragments, and a series of user-defined atomic constraints (C, H, N, O, P, F, S, Cl, I, and Br) (level 4). We then used the package MetFragR[64] to compare measured $MS^2$ fragmentation patterns with *in silico* fragmentation patterns of all chemicals in the PubChem[63] online database with a molecular formula that matches the predicted molecular formula. In cases where an unequivocal molecular formula could not be assigned to a nontarget MS feature (*i.e.*, more than one molecular formula assigned with a similar score), we used its accurate mass to search the PubChem database instead. The resulting list of candidate chemical structures (level 3) was ranked based on the weighted scoring of six factors: fragment score (0.30); Metfusion score[122] (0.30); number of PubChem references (0.05); number of PubChem patents (0.05); RT score (0.15); and presence in SusDat, the merged NORMAN suspect list[123] (0.15). All scores are normalized (0 – 1) to the top ranked candidate for each individual scoring metric. These types of scoring factors have been used in previous studies[55,64], but the weighting factors were optimized

to maximize the correct annotations of target micropollutants among the samples collected from the first few weeks of the FCMS. The most plausible structure(s) was selected based on the scoring and other chemical information (name, use-class, clustering with target micropollutant) (level 2P). Finally, the fragmentation patterns of the selected candidate structures were compared to online mass spectral libraries including MassBank of North America (MoNA)[111] and mzCloud[109] (level 2L) or an in-house library that included major diagnostic $MS^2$ fragments and RTs for over 600 micropollutants (level 1L). If available, a pure reference standard was obtained for unambiguous confirmation (level 1). The complete pipeline to assign chemical structures to the prioritized nontarget MS features is available for download at github.com/cmc493.

## 3.4 Results and Discussion

### *3.4.1 Peak picking and profile generation*

Our sample set consisted of 361 daily composite samples (4 samples were lost throughout the year) and 51 field blanks. Our optimized peak picking workflow identified 18.3 million fully-resolved chromatographic peaks among the 361 samples, some of which represent the same constituent identified in multiple samples. The workflow also includes a replicate filter, which excludes picked peaks that are not present in all three of the triplicate sample injections; the replicate filter excluded approximately 25% of the picked peaks, leaving 13.8 million peaks to be included in the resulting 300,309 profiles of MS features.

### *3.4.2 MS feature profiles of target micropollutants*

Our optimized peak picking and profile generation workflow resulted in 31 target micropollutant annotations among the 162 target micropollutants included in the study. Although some of the target micropollutants were sporadically detected in few samples and did not generate

a continuous temporal profile, many target micropollutants were more ubiquitously present and generated temporal profiles that provide new insights into temporal dynamics of micropollutant abundance in small streams. For example, some frequently detected target micropollutants that are often associated with STP outfalls (*e.g.*, desvenlafaxine, fexofenadine, and lamotrigine) had relatively high normalized intensities during periods of low streamflow and relatively low normalized intensities during periods of high streamflow, as demonstrated in **Figure 3.1A** for desvenlafaxine. This apparent negative association between micropollutant abundance and streamflow suggests a continuous loading into Fall Creek and subsequent dilution during periods of wet weather. Conversely, the abundance of other frequently detected target micropollutants that are often associated with agricultural activities (*e.g.*, atrazine, metolachlor, simazine) exhibited a strong positive association with streamflow during the agricultural season (approximately June through August) and had low to no abundance throughout the remainder of the study period, as demonstrated in **Figure 3.1B** for atrazine. These data provide evidence that micropollutants associated with agricultural activities can be mobilized during precipitation events and their abundance increases even as the amount of water in Fall Creek increases, reflecting significantly increased mass loading during runoff events. Whereas these general types of associations between streamflow and the abundance of STP-derived[124–126] and agriculture-derived[35,100,127] micropollutants have been previously observed, the data reported here offer the first description of the dynamics of micropollutant abundance in daily composite samples collected from a small stream over an annual period. These data have value for evaluating the dynamics of chemical exposure in small streams and could inform decision making at water utilities using small streams as a drinking water source. The temporal profiles of the other 29 target micropollutants are plotted along with streamflow in **Figures B2 – B30**.

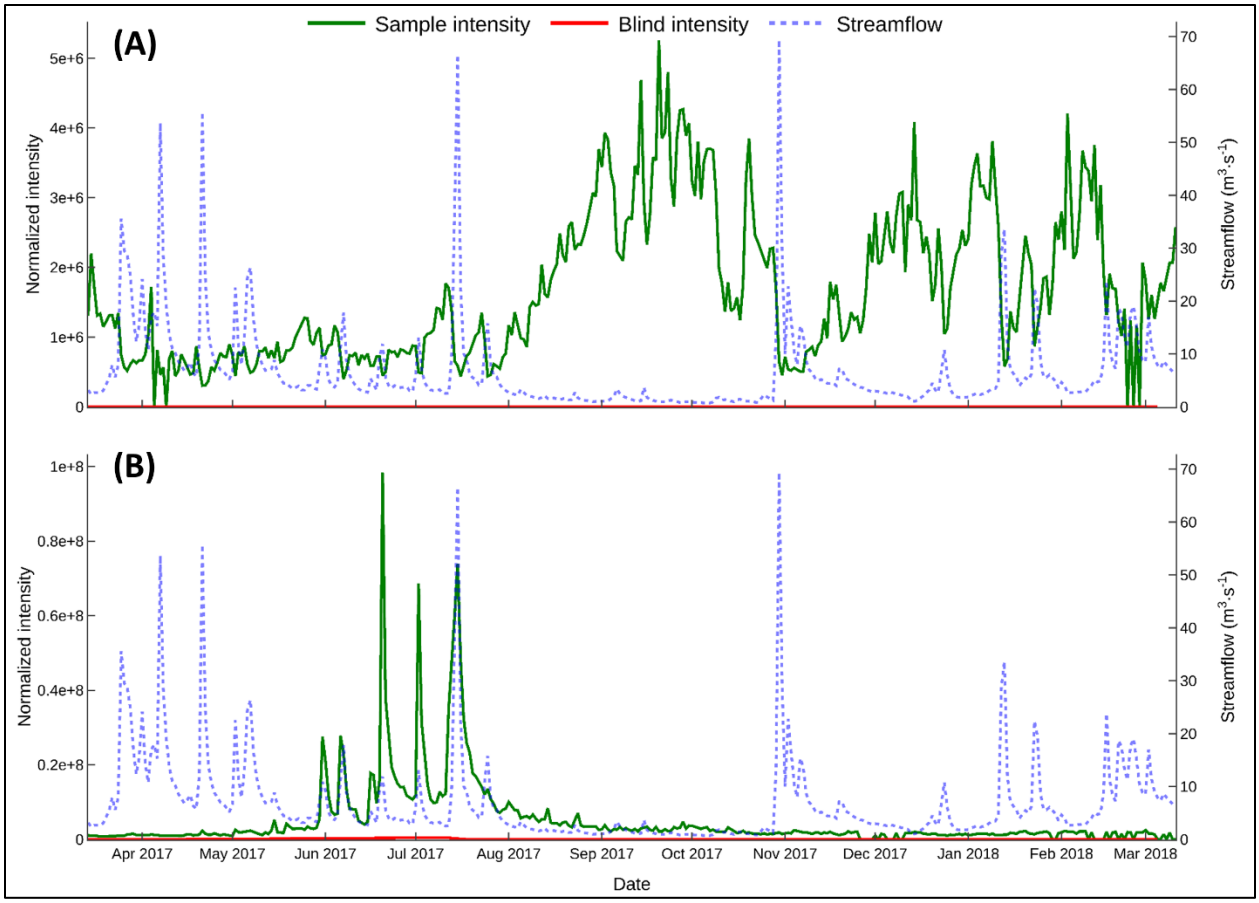



**Figure 3.1:** Micropollutant temporal trend profiles and streamflow for target micropollutants (A) STP-derived desvenlafaxine and (B) agriculture-derived atrazine.

### *3.4.3 Filtering and clustering of MS feature profiles*

As a means to prioritize profiles of nontarget MS features for structural elucidation, we applied a series of data reduction filters to exclude profiles that do not meet certain quality control metrics. We first assigned conservative thresholds for sample-to-blind intensity ratio (≥ 10), median trend intensity (≥ $10^5$), and RT (≥ 6.5 min) based on a meta-analysis of these properties in the full profile list (**Figure B31**). The sample-to-blind ratio filter excluded 28% of the profiles that were present in both the samples and field blanks at median sample-to-blind intensity ratios less than 10. The mean trend intensity filter excluded 45% of the remaining profiles with mean trend intensities less than $10^5$. The RT filter excluded 6% of the remaining profiles containing picked

peaks that eluted near the solvent front and had poor chromatographic retention and peak shape; we note that some very polar micropollutants (*e.g.*, metformin) may have been excluded from the profile list after applying this filter. We next excluded approximately 1% of the remaining profiles associated with lower order isotopologues and adducts; it is worth noting that most of the profiles excluded here would not have been excluded during any other filtering or prioritization step and would therefore have been included in the final profile list. Finally, as a means to prioritize the remaining profiles containing a continuous temporal profile for at least 3% of the study period, we excluded all MS feature profiles that did not contain at least 30 total detections and at least 10 consecutive detections among the annual daily samples. The final list contains 1,981 filtered MS feature profiles (0.66% of the total number of profiles), including 18 of the 31 target micropollutants that were originally annotated; most of the target micropollutants that were excluded from the final list were removed during the final filtering step because they were only sporadically detected throughout the study period. More details on our overall data reduction approach are provided in **Appendix B**.

We then grouped the 1,981 filtered MS feature profiles by means of HCA. The resulting dendrogram is presented in **Figure 3.2** which highlights the locations of the 18 target micropollutants. We hypothesized that the highly resolved temporal profiles of the target micropollutants could be used as a means to prioritize nontarget MS features for structural elucidation; we expect that micropollutants represented by closely clustered MS features will have similar sources, use-classes, and fate and transport properties within the watershed.[21,98] For example, nontarget MS features that are clustered closely to atrazine, metolachlor, and simazine (**Figure 3.2**) are likely to be micropollutants that are also related to agricultural activities and have similar transport behavior. We prioritized 115 nontarget MS feature profiles (6% of the filtered

profiles) that were clustered within or adjacent to localized clusters containing target micropollutants. To the best of our knowledge, this is the first study to prioritize nontarget MS features based on highly resolved temporal profiles and their relationship with respect to the profiles of target micropollutants. Inferring the use-classes or sources of the nontarget micropollutants aids in the identification of the unknown chemical structures by narrowing the breadth of candidate structures.

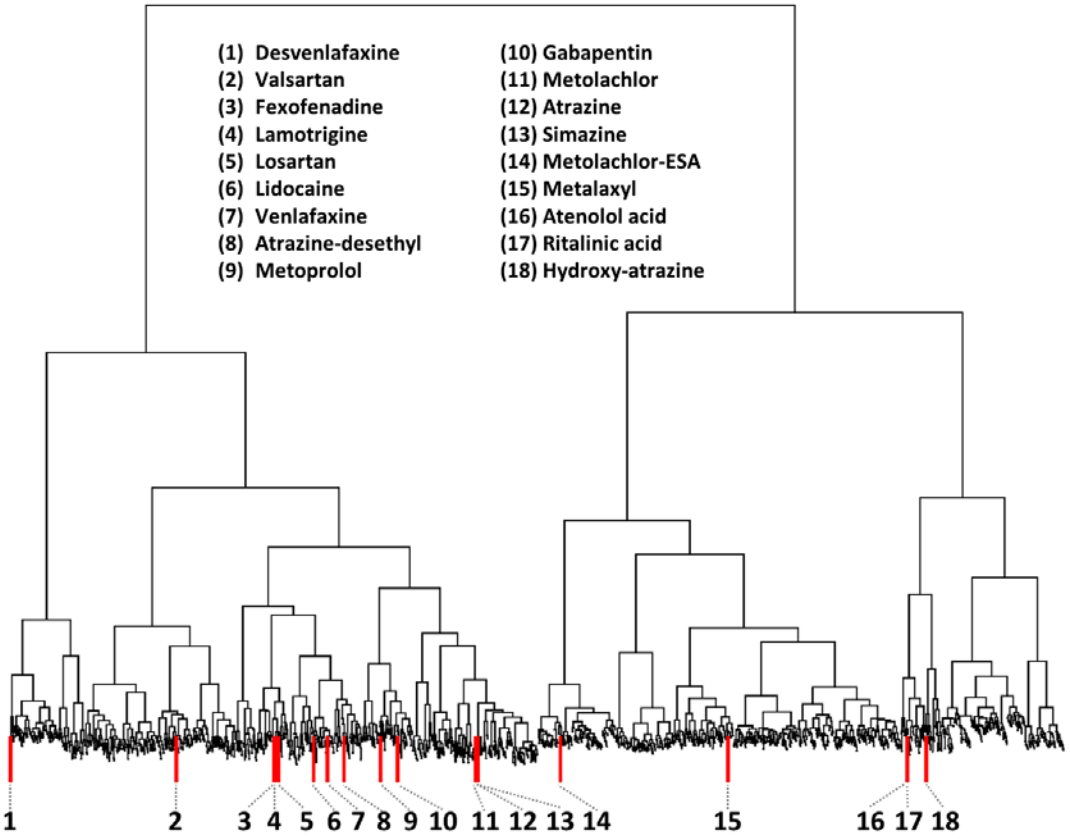



**Figure 3.2:** Dendrogram of filtered profiles (n = 1,981). Marked profiles indicate target micropollutants.

### *3.4.4 Structural elucidation of nontarget MS features*

Each of the nontarget MS features included in the dendrogram are described by their average *m/z* and RT among all of the samples and are considered to be exact masses of interest (level 5). To facilitate the structural elucidation of each of the prioritized nontarget MS feature profiles, we developed and applied an automated pipeline to assign chemical structures using a

series of self-written R scripts and publically available R packages. Similar approaches have been described for structural elucidation of nontarget micropollutants in other environments,[56,104] though our pipeline offers high-throughput data processing and structural elucidation, is fully transparent and customizable, and has been made publically available for free use within the R environment. Our approach allowed us to identify 59 nontarget micropollutants with an increased level of confidence (in addition to the 31 target micropollutants previously identified), including 14 confirmed nontarget micropollutant structures (level 1 or 1L), 8 probable structures (level 2L or 2P), 9 tentative candidates (level 3), and 28 unequivocal molecular formulae (level 4). The nontarget micropollutants identified with a confidence of level 3 or higher are listed in **Table 3.1**, along with the target micropollutants with which they clustered in the dendrogram.

An example of how the pipeline is used to elucidate the structure of a nontarget MS feature (NT242, $m/z_{avg}$ = 242.1306, $RT_{avg}$ = 12.2 min) is provided in **Figure 3.3**. First, a nontarget MS feature profile is selected based on its proximity to a target micropollutant in the dendrogram and its temporal profile is displayed (**Figure 3.3A**); NT242 clustered with desvenlafaxine suggesting it may be a STP-derived micropollutant. Then, the sample with the highest ILIS normalized intensity (October 5) is selected and the associated MS and $MS^2$ data are automatically loaded into R from a local .mzXML file to display the extracted ion chromatogram (**Figure 3.3B**). The RT of the nontarget MS feature is used to extract the associated $MS^1$ scan to display the isotopic signature (**Figure 3.3C**) and the nearest $MS^2$ scan to display the fragmentation pattern (**Figure 3.3D**). The average *m/z* of the nontarget MS feature is then automatically used to predict the molecular formula within a 5 ppm mass deviation (Δm), and the $MS^1$ and $MS^2$ data are used to score the prediction.[121] The top four scored molecular formulae for NT24 are provided in **Figure 3.3E**; the top scored molecular formula is $C_{13}H_{20}NOCl$ (Δm = 0.10 ppm).

**Table 3.1:** List of detected target and nontarget micropollutants (level 3 or above) with associated molecular formula or CAS number (if available) and identification confidence level.

| Name | Molecular Formula or CAS No. | Level | Name | Molecular Formula or CAS No. | Level |
|---|---|---|---|---|---|
| **Desvenlafaxine** | 93413-62-8 | 1 | **Simazine** | 122-34-9 | 1 |
| Rac-threo-dihydrobupropion | 99102-04-2 | 1 | Esculentic Acid | 464-92-6 | 2P |
| Gabapentin-lactam | 64744-50-9 | 1 | 2-methylthio-benzothiazole | 615-22-5 | 1 |
| 2-mercapto-benzothiazole | 149-30-4 | 1 | NT270 (mercapto-benzothiazole sub-structure) | $C_{11}H_{11}NO_3S_2$ | 3 |
| NT240 (benzothiazolone sub-structure) | $C_{11}H_{13}NO_3S$ | 3 | NT344 (mercapto-benzothiazole sub-structure) | $C_{15}H_{21}NO_2S_3$ | 3 |
| **Valsartan** | 137862-53-4 | 1 | **Metolachlor-ESA** | 171118-09-5 | 1 |
| 8-benzyloxy-2,3-dihydro-1,4-benzodioxin-5-carboxylic acid | 69114-85-8 | 2P | Propazine-2-hydroxy | 7374-53-0 | 1L |
| 4-butoxy-N,N-bis (2-ethoxyethyl) benzenesulfonamide | $C_{18}H_{31}NO_5S$ | 2P | **Metalaxyl** | 57837-19-1 | 1 |
| **Fexofenadine** | 83799-24-0 | 1 | Metolachlor-OXA | 152019-73-3 | 1 |
| **Lamotrigine** | 84057-84-1 | 1 | Nootkatone | 4674-50-4 | 2L |
| NT264 (guanidine sub-structure) | $C_{16}H_{29}N_3$ | 3 | NT212 (hydroxy-s-triazine sub-structure) | $C_9H_{17}N_5O$ | 3 |
| **Losartan** | 114798-26-4 | 1 | 2-amino-benzothiazole | 136-95-8 | 1 |
| **Lidocaine** | 137-58-6 | 1 | **Atenolol acid** | 56392-14-4 | 1 |
| Lauric diethanolamide | 120-40-1 | 2P | **Ritalinic acid** | 19395-41-6 | 1 |
| Methyl diethyl-dithiocarbamate | 686-07-7 | 1 | **Hydroxy-atrazine** | 2163-68-0 | 1 |
| NT180 (thiocarbamate sub-structure) | $C_6H_{13}NOS_2$ | 3 | 2,2'-Dithiobis-benzothiazole | 120-78-5 | 1 |
| NT148 (thiocarbamate sub-structure) | $C_6H_{13}NOS$ | 3 | 1,2-dihydro-2,2,4-trimethylquinoline | 147-47-7 | 1 |
| (4S)-4-Amino-5-(dibutylamino)-5-oxopentanoic acid | $C_{13}H_{26}N_2O_3$ | 2P | Triethyl phosphate | 78-40-0 | 1 |
| **Venlafaxine** | 93413-69-5 | 1 | *Targets removed during profile filtering* | | |
| Dimethyl phthalate | 131-11-3 | 1 | **1-Methyl-benzotriazole** | 136-85-6 | 1 |
| **Atrazine-desethyl** | 6190-65-4 | 1 | **Benzotriazole** | 95-14-7 | 1 |
| Triphenyl phosphate | 115-86-6 | 1 | **Bupropion** | 34841-39-9 | 1 |
| NT343 (#-hydroxyphenyl diphenyl phosphate) | $C_{18}H_{15}O_5P$ | 3 | **Caffeine** | 58-08-2 | 1 |
| NT222 (#-anilinoquinazoline) | $C_{14}H_{11}N_3$ | 3 | **Diethyl phthalate** | 84-66-2 | 1 |
| **Metoprolol** | 37350-58-6 | 1 | **Fluconazole** | 86386-73-4 | 1 |
| **Gabapentin** | 60142-96-3 | 1 | **Irbesartan** | 138402-11-6 | 1 |
| **Metolachlor** | 51218-45-2 | 1 | **Methocarbamol** | 532-03-6 | 1 |
| Alachlor-OXA | 171262-17-2 | 1 | **Prometon** | 1610-18-0 | 1 |
| Metolachlor-2-hydroxy | 131068-72-9 | 2P | **Propazine** | 139-40-2 | 1 |
| **Atrazine** | 1912-24-9 | 1 | **Sitagliptin** | 486460-32-6 | 1 |
| Nuciferine N-oxide | 104385-30-0 | 2P | **Trimethoprim** | 738-70-5 | 1 |
| | | | **Warfarin** | 2610-86-8 | 1 |

Notes: The table is organized to present micropollutants in the order in which they were clustered (from left to right) in the dendrogram provided in **Figure 3.2**; **Figures B32 – B90** describe the temporal trends and the MS information for each nontarget micropollutant; bolded names represent target micropollutants.

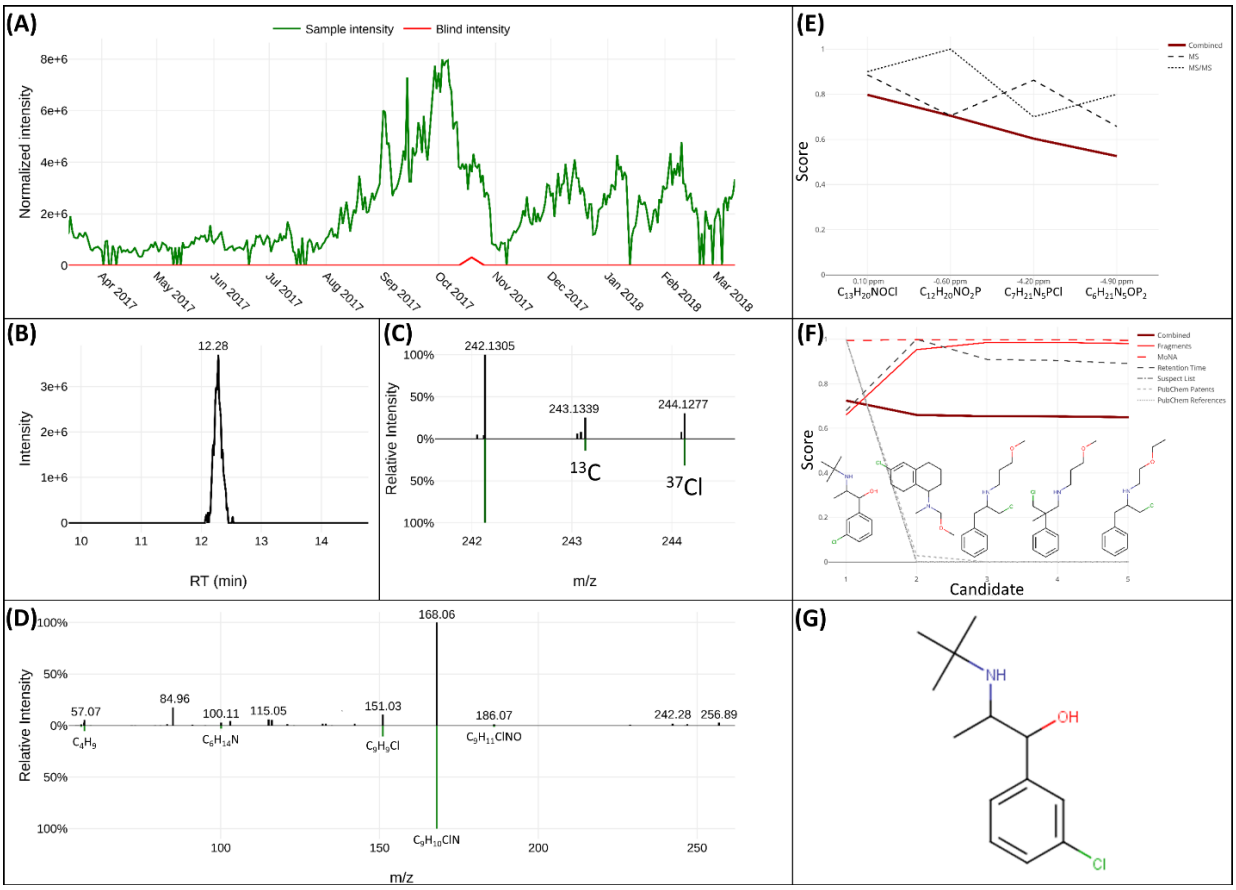



**Figure 3.3:** Identification of NT242 following the structure elucidation workflow: (A) temporal trend profile; (B) extracted ion chromatogram; (C) measured MS spectra (top) and theoretical MS spectra (bottom); (D) measured $MS^2$ fragmentation pattern (top) and *in silico* $MS^2$ annotated fragments (bottom); (E) molecular formula prediction; (F) candidate structure ranking; and (G) confirmed structure. NT242 was confirmed as rac threo-dihydrobupropion (level 1) using an authentic reference standard which matched the $RT_{avg}$ of 12.2 min, the MS spectra ($m/z_{avg}$ = 242.1306 for $[M+H]^+$, $\Delta m$ = 0.13 ppm), the theoretical abundance (13%) of the $^{13}C$ monoisotopic mass, the theoretical abundance (32%) of the $^{37}Cl$ monoisotopic mass, and the $MS^2$ fragments ($m/z$ = 57.07, 151.03, and 168.06).

Next, the pipeline implements MetFrag and uses the top scored molecular formula to obtain candidate structures from the PubChem online database, which are further filtered to remove non-connected compounds (*e.g.*, salts) and lower-order isotopes.[64] The final list of candidate structures is then subjected to *in silico* fragmentation[64] and ranked based on a weighted score that we optimized to maximize the scoring of the 31 target micropollutants previously identified. The fragment score (0.30) is determined by how well the measured $MS^2$ fragments are explained by

the *in silico* $MS^2$ fragments. The Metfusion score (0.30) is determined by how well the measured $MS^2$ fragments match with fragments of similar structures in the MoNA spectral library.[122] The PubChem score is determined by the total number of PubChem references (0.05) and patents (0.05) for a candidate structure, which is a metric that can significantly improve the accuracy of nontarget structure elucidation.[64,128–130] The RT score (0.15) is determined by how well the expected RT of a candidate structure aligns with the RT of the nontarget MS feature. We used the data from our 162 target micropollutants to develop a linear relationship between the octanol-water partition coefficient (logP) and measured RT. The expected logP of the nontarget MS feature is predicted and compared to the estimated logP of the candidate structure for scoring.[64,131] The suspect score (0.15) is determined based on whether the candidate structure is included in SusDat, the merged NORMAN suspect list;[123] we reasoned that candidate structures that have been previously detected or suspected as a water contaminants should receive a higher score. The top five candidate structures for NT242 and their scoring are provided in **Figure 3.3F**. More details on our structural elucidation pipeline and the weighting factors are provided in **Appendix B**.

The most plausible candidate structure(s) based on scoring and other chemical information (name, use-class, clustering with target micropollutant) is selected from the ranked MetFrag candidate list. The top scored candidate structure for NT242 is rac threo-dihydrobupropion (**Figure 3.3G**), a transformation product of the pharmaceutical bupropion. This is a plausible identification because the temporal profile of NT242 clustered in close proximity to the pharmaceutical desvenlafaxine and bupropion was one of the 31 target micropollutants identified in this study. The acquired MS data for plausible candidate structures can be compared with online and in-house mass spectral libraries, though no MS data for rac threo-dihydrobupropion was found in any library. Therefore, an authentic reference standard was acquired for

rac threo-dihydrobupropion and the data acquired for the authentic standard matched the data acquired for NT242, confirming the level 1 identification. To the best of our knowledge, rac threo-dihydrobupropion has not been previously reported as a micropollutant. It is interesting to note that we measured a continuous temporal profile of rac threo-dihydrobupropion in Fall Creek, but only an intermittent profile for bupropion; if one assumes that the two micropollutants have the same sources, this suggests that there may be differential fate and transformation behavior of the two micropollutants. Temporal profiles and the analytical data supporting the identification of the 59 nontarget micropollutants are provided in **Figures B32 – B90**.

### *3.4.5 Other nontarget micropollutants*

Our workflow resulted in the identification of 59 nontarget micropollutants in Fall Creek, with varying levels of confidence. Several notable classes of nontarget micropollutants were identified. First, six of the nontarget micropollutants could be classified as rubber vulcanization accelerators, which are rarely reported as environmental micropollutants. For example, 2-mercaptobenzothiazole (MBT, level 1) has been reported in industrial wastewaters and stormwater, but to the best of our knowledge, has not been reported in surface water systems as it is easily transformed in the environment.[132–134] We also identified putative transformation products of MBT including 2,2'-dithiobisbenzothiazole (DTBT, level 1) and 2-methylthiobenzothiazole (MTBT, level 1), along with other compounds with MBT substructures (NT270 and NT344, level 3) and a benzothiazoline substructure (NT240, level 3). MTBT has been reported in wastewater,[135,136] and DTBT has rarely been detected in surface water.[137] The source of rubber vulcanization accelerators in the Fall Creek watershed is unknown, though we speculate that vehicular tire wear and subsequent transport in road runoff may be the source.[76,138] These findings

are particularly relevant as a recent study has suggested that micropollutants derived from tire wear may be linked to high toxicity events in small streams.[21]

Next, five micropollutants that could be classified as plasticizers or flame retardants were identified in Fall Creek including diethyl phthalate (level 1, target micropollutant), dimethyl phthalate (level 1), triethyl phosphate (level 1), triphenyl phosphate (level 1), and hydroxyphenyl diphenyl phosphate (level 3). Phthalate esters, including diethyl phthalate and dimethyl phthalate, have been previously reported as environmental micropollutants and are associated with multiple sources including urban runoff and wastewater effluents.[139] Likewise, organophosphorous plasticizers and flame retardants such as triethyl phosphate and triphenyl phosphate have also been reported as environmental micropollutants in surface waters and are mainly associated with wastewater sources.[140,141] To the best of our knowledge, hydroxyphenyl diphenyl phosphate is reported here for the first time and is likely a transformation product or manufacturing impurity of triphenyl phosphate, though its source cannot be determined from our data.

A number of nontarget pesticides and pesticide transformation products were also identified with temporal profiles that fall into three distinct clusters. First, alachlor-OXA (level 1) and metolachlor-2-hydroxy (level 2P) were identified based on their clustering near atrazine (level 1, target micropollutant), metolachlor (level 1, target micropollutant), and simazine (level 1, target micropollutant). This cluster of pesticides exhibits a strong positive association with streamflow during the agricultural season (approximately June through August) and has low to no abundance throughout the remainder of the study period. Second, another cluster of pesticide transformation products were more ubiquitously present throughout the study period and exhibited little association with streamflow. These include atrazine-hydroxy (level 1, target micropollutant), metolachlor-ESA (level 1, target micropollutant), and propazine-2-hydroxy (level 1L). We suspect

that the source of these transformation products is from groundwater recharge.[61] Finally, a third cluster of pesticides and pesticide transformation products exhibited a positive association with streamflow throughout the study period, suggesting persistence and continued mobilization throughout the year. These include metalaxyl (level 1, target micropollutant), metolachlor-OXA (level 1), and a hydroxy-*s*-triazine-containing compound (NT212, level 3). These data allow us to identify three distinct clusters of agriculture-derived micropollutants that provide insights on their relative persistence and fate throughout an annual period. Our results indicate that different pesticides and pesticide transformation products have varying concentration-discharge dependencies throughout the year, which is an important finding for exposure assessment and source control.

Finally, our sampling strategy resulted in samples being stored for varying amounts of time (between 0 and 6 days) inside the automatic sampler. Whereas most micropollutants were stable during this storage, some exhibited evidence of degradation during storage. This manifests as a periodic saw tooth pattern in the temporal profile, as shown in **Figure B47**. For example, several micropollutants with thiocarbamate substructures including methyl diethyldithiocarbamate (MeDDC, level 1), NT180 (level 3), and NT148 (level 3) exhibited this behavior. MeDDC is a human metabolite of disulfiram, a pharmaceutical used to treat alcoholism; disulfiram was not detected in a retrospective screening of the high-resolution mass spectral acquisition. To the best of our knowledge, MeDDC has not been reported as a micropollutant. Thio- and dithiocarbamates are found in many fungicides and their instability during storage suggests limited persistence in Fall Creek.

### *3.4.6 Insights and environmental implications*

A primary goal of FCMS is to enable a continuous and comprehensive characterization of organic micropollutant occurrence in a drinking water source. The data presented here represent results from the first year of samples collected from the FCMS and describe our efforts to establish high-throughput sample preparation, sample analysis, and post-acquisition data processing workflows to meet this goal. Whereas we used temporal profiles to prioritize nontarget MS features for structural elucidation, our data processing workflow and pipeline can be used to elucidate chemical structures of any nontarget MS feature in high-resolution mass spectral acquisitions from any type of sample including dust, soil, sediment, blood, serum, or wastewater. Our data reveal new insights on the dynamics of micropollutant occurrence in a small stream. For example, important and distinct concentration-discharge relationships were noted for both STP-derived and agriculture-derived micropollutants, though these general relationships alone are insufficient to explain the temporal dynamics of specific micropollutants. Concentration-discharge relationships are expected to be masked or equalized in larger surface water systems, but small streams are clearly vulnerable to hydrological events within the watershed and further research is warranted to study associations among temporal micropollutant profiles and various watershed features. This work is a first step towards improving our ability to characterize the dynamics of exposure risk in small streams and to predict peak events while simultaneously considering multiple contaminants.

# CHAPTER 4 – Fall Creek Monitoring Station: Using Environmental Covariates to Predict Micropollutant Dynamics and Peak Events in Surface Water Systems[c]


## Abstract

This research aimed to further our understanding of how environmental processes control micropollutant dynamics in surface water systems as a means to predict peak events and inform intermittent sampling strategies. We characterized micropollutant abundance in daily composite samples from the Fall Creek Monitoring Station (FCMS) over 18 months. These data were compiled alongside environmental covariates, including daily measurements of weather, hydrology, and water quality parameters, to generate a novel dataset with high temporal resolution. We evaluated the temporal trends of several representative micropollutants, along with cumulative metrics of overall micropollutant contamination, by means of multivariable analyses to determine which combination of covariates best predicts micropollutant dynamics and peak events. Peak events of agriculture-derived micropollutants were best predicted by positive associations with turbidity and $UV_{254}$ absorbance, and negative associations with baseflow index. Peak events of wastewater-derived micropollutants were best predicted by positive associations with alkalinity and negative associations with streamflow rate. We demonstrate that these predictors can be used to inform intermittent sampling strategies aimed at capturing peak events, and generalize the approach so that it could be applied in other watersheds. Finally, we demonstrate how our approach can be used to contextualize micropollutant data derived from infrequent grab samples.

## 4.1 Introduction

Micropollutants are anthropogenic organic chemicals that are present in natural and engineered water systems at low concentrations.[1] Micropollutant monitoring studies carried out over the past two decades have resulted in the identification of hundreds of micropollutants in surface water systems around the world.[2,3,32] These historic data reveal two important observations on the global micropollutant situation. First, many types of micropollutants are inherently bioactive and reported micropollutant concentrations in the environment raise concerns about health effects on exposed ecosystems[12,14] and human populations.[13] Second, the occurrence and concentration of micropollutants in aquatic systems is dynamic and can exhibit significant spatiotemporal variability.[35,98,142] This noted spatiotemporal variability is an essential consideration when developing micropollutant monitoring strategies in the context of environmental risk assessment;[38] intermittent grab samples that target a finite set of micropollutants are likely to significantly underestimate the chemical risk in that system.[33]

The dynamic nature of micropollutant occurrence and abundance in surface water systems is driven by variable micropollutant emissions and the sum of all environmental processes that determine micropollutant fate and transport.[143] If the emissions of a particular class of micropollutant are well-defined, then the occurrence and concentration of any micropollutant in that class might be predicted by a set of environmental covariates that control the fate and transport of that micropollutant. For example, the concentrations of twelve wastewater-derived micropollutants measured in monthly grab samples were associated with dissolved oxygen levels, solar radiation, pH, chlorophyll a concentration, and antecedent dry period.[44] The concentrations of two agriculture-derived micropollutants measured in weekly samples collected in 100 small streams revealed that atrazine concentrations were associated with baseflow index while

glyphosate concentrations were not, an observation attributed to differing sorption affinities.[144] Concentrations of benzotriazoles measured in grab samples collected across three watersheds were linked to upstream road density, precipitation, and snowmelt.[145] Finally, a study examining several types of micropollutants in intermittent grab samples from a mixed-use watershed found higher concentrations of agriculture-derived micropollutants during high flow events, and higher concentrations of wastewater-derived micropollutants during low flow events.[35] Together, these studies demonstrate that examining micropollutant occurrence and concentration data coupled with environmental covariates can lead to the identification of predictors of micropollutant dynamics at watershed scales. However, the limited scope of these and similar datasets have not allowed for generalizable conclusions that can be used to inform micropollutant sampling strategies that are optimized for environmental risk assessment.

The primary goal of this study was to improve our fundamental understanding of the drivers of micropollutant dynamics in surface water systems. The objectives of the study were to: (i) identify environmental covariates that can be used to predict the temporal concentration profiles of representative micropollutants and the temporal profiles of cumulative metrics of overall micropollutant contamination; (ii) identify environmental covariates that are predictive of peak events among the selected response variables; and (iii) explore the development of informed sampling strategies as a means to improve the accuracy of environmental risk assessments performed with datasets obtained from intermittent sampling. We used data from the Fall Creek Monitoring Station (FCMS) and a set of environmental covariates (*i.e.*, weather, hydrological, and water quality parameters) to build multivariable models that explain the dynamics of measured micropollutant concentrations in Fall Creek, a small stream that serves as a drinking water source for over 30,000 people. We previously reported on a novel analytical and post-acquisition data

processing workflow to comprehensively characterize micropollutant occurrence in daily surface water samples obtained from the FCMS.[146] Here, we used 18 months of data from the FCMS to build temporal micropollutant profiles (MPPs) for the concentrations of eight representative micropollutants along with temporal profiles for cumulative metrics of overall micropollutant contamination. We identified environmental covariates that are predictive of dynamic micropollutant concentrations and peak events in Fall Creek. These results were then used to propose informed intermittent sampling strategies that can improve the accuracy of environmental risk assessments.

## 4.2 Material and Methods

### *4.2.1 Modelling framework*

The modelling framework is described in **Figure 4.1** and consists of three steps: (1) data collection, (2) data processing, and (3) multivariable regression. All analyses were conducted with the R Statistical Software v3.3.[120]

### *4.2.2 Data Collection – micropollutant profiles*

The FCMS is located at the raw water intake of the Cornell Water Filtration Plant (CWFP) and draws water from Fall Creek, which had an average discharge of 6 $m^3 \cdot s^{-1}$ (median 3.5 $m^3 \cdot s^{-1}$, range 0.6 – 70 $m^3 \cdot s^{-1}$) during the study period.[79] Details on the FCMS including sample collection, sample preparation and analysis, and post-acquisition data processing are provided elsewhere.[146] Briefly, we collected daily, time-proportional composite samples from the FCMS for 18 months (552 days from March 2017 through September 2018); we extended our previously reported FCMS sampling campaign by 6 months to collect samples through an additional agricultural growing season. Flowrates during the study period are representative of long term trends and flow duration

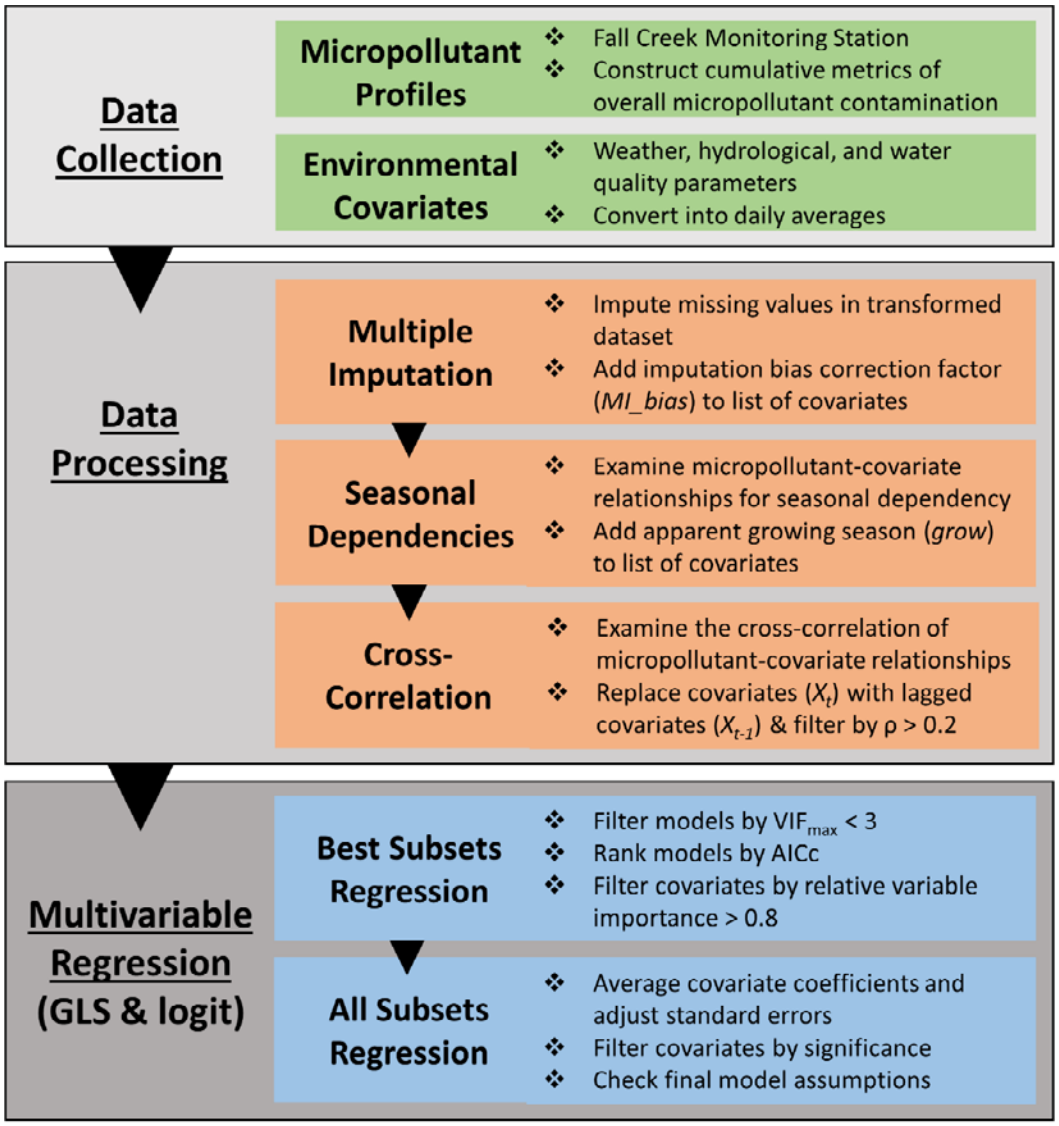



**Figure 4.1:** The modelling framework consisted of three steps: (1) data collection, (2) data processing, and (3) multivariable regression. Micropollutant profiles were obtained using previously established methods.[146] The modelling framework was applied using generalized least squares (GLS) and logistic (logit) regression techniques.

curves are provided in **Figure C1** of **Appendix C**.[147] Full-scan mass spectra were acquired for each sample using large volume injection and high-performance liquid chromatography coupled with high-resolution mass spectrometry (QExactive hybrid quadrupole orbitrap, Thermo Fisher Scientific). Our previous study reported temporal profiles on the relative abundance of 31 target and 59 nontarget micropollutants with varying levels of confidence over 12 months of operation

of the FCMS.[61] This study examines temporal concentration profiles for micropollutants that were reported with level 1 confidence (*i.e.*, unequivocally confirmed structure) over an 18 month period. Micropollutant concentrations were estimated using authentic reference standards and an external calibration. Quantification was based on isotope labeled internal standard-normalized intensities calculated using enviMass[53] and 1/x weighted linear least-squares regression.

We aimed to explore the drivers of micropollutant fate and transport for both agriculture and wastewater-derived micropollutants. Therefore, we selected atrazine (ATR) and metolachlor (MET) as representative agriculture-derived micropollutants along with their transformation products atrazine-desethyl (ATR.d), atrazine-2-hydroxy (ATR.h), and metolachlor-OXA (MET.o). We selected desvenlafaxine (DES), gabapentin-lactam (GAB.l), and rac-threo-dihydrobupropion (DHB) as representative wastewater-derived micropollutants. We also built temporal profiles that describe cumulative metrics of overall micropollutant contamination. The cumulative normalized abundance (CUMA) profile describes the normalized cumulative abundance of all 90 previously reported micropollutants (*i.e.*, temporal abundance profiles of all 90 detected micropollutants were normalized to the maximum abundance in each profile and then summed together). The cumulative estimated concentration (CUMC) profile describes the cumulative estimated concentrations of the 42 micropollutants that we previously identified with level 1 confidence. Because some micropollutants had estimated concentrations that exceeded the range of our calibration, we also defined the CUMC2 profile, which describes the cumulative estimated concentrations of the 32 micropollutants that were consistently measured within the limits of our calibration over the duration of the study period. The cumulative estimated risk (CUMT) profile describes the estimated cumulative risk of the 17 micropollutants that were identified with level 1 confidence, consistently met our quality control criteria with respect to field

blanks and limits of our calibration, and for which toxicity data was available within the United States Environmental Protection Agency's (USEPA) ToxCast database.[148] Cumulative risk was calculated by summing the exposure-activity ratio (EAR) profiles of each of the micropollutants using the R package *toxeval*[149] as previously described.[13,150–152]

### *4.2.3 Data Collection – environmental covariates*

We collected a set of environmental covariates that have been previously reported or are expected to contribute to micropollutant dynamics at watershed scales. The environmental covariates are summarized in **Table 4.1** and detailed in **Appendix C**. Spatially variable environmental covariates (*i.e.*, weather data) were converted into daily averages or totals across the Fall Creek watershed. Hydrological and water quality data were converted into daily averages for Fall Creek. Weather data and streamflow rates were obtained from open sources.[79,153–155] Sewage treatment plant (STP) discharge rates and water quality data were obtained from the respective STP and the CWFP.

### *4.2.4 Data Processing – multiple imputation*

Multiple imputation was applied to generate complete datasets by filling missing environmental covariate and micropollutant data using previously described methods.[42] Briefly, we used the R package *MICE*[156] and the predicted mean matching imputation method to generate five possible values for each missing value and consequently five complete imputed datasets. Prior to multiple imputation, the environmental covariates (excluding the binary parameters) were centered and scaled,[157] and several variables were transformed to increase the linearity of the response and to transform the data to an approximately normal distribution.[158,159] Micropollutant concentrations, *flow*, *turb*, *uv*, *stp*, and *pstp* were log transformed and temperature covariates were squared. First-order lags and leads for each variable were added to the dataset to improve the

**Table 4.1:** Information of the selected environmental covariates.

| Environmetnal covariate[a] | Abbreviation | Units | Data source (n = number of weather stations)[b] or Formula |
|---|---|---|---|
| ***Weather*** | | | |
| Precipitation (total) | $prcp$ | *mm* | NOAA & NRCC (n = 56) |
| Precipitation index | $prcp01$ | - | $prcp01_t = \begin{cases} 1, prcp_t \geq 1 \\ 0, prcp_t < 1 \end{cases}$ |
| Antecedent dry period | $adp$ | *days* | $ADP_t =$ days since $prcp_t > 1$ |
| Temperature | $tavg$ | *°C* | NOAA & NRCC (n = 8) |
| Temperature (maximum) | $tmax$ | *°C* | NOAA & NRCC (n = 21) |
| Temperature (minimum) | $tmin$ | *°C* | NOAA & NRCC (n = 21) |
| Snowfall | $snow$ | *mm* | NOAA (n = 20) |
| Snow depth | $snwd$ | *mm* | NOAA (n = 15) |
| Snow melt | $snmt$ | *mm* | $snmt = snwd_t - snwd_{t-1}$ |
| Wind speed | $wind$ | *kph* | NOAA & NRCC (n = 10) |
| Leaf wetness (total) | $lwet$ | *min* | NRCC (n = 4) |
| Solar radiation (total) | $srad$ | *langleys* | NRCC (n = 4) |
| Relative humidity | $rhum$ | *%* | NRCC (n = 6) |
| ***Hydrology*** | | | |
| Streamflow rate | $flow$ | $m^3 \cdot s^{-1}$ | USGS |
| Baseflow index[c] | $bfi$ | - | $bfi_t = \frac{\text{baseflow}(flow)_t}{flow_t}$ |
| STP discharge | $stp$ | $m^3 \cdot s^{-1}$ | Local municipality (in-line) |
| Sewage proportion | $pstp$ | *%* | $pstp_t = \frac{stp_t}{flow_t}$ |
| ***Water Quality*** | | | |
| pH | $pH$ | - | CU WFP (in-line) |
| Turbidity | $turb$ | *NTU* | CU WFP (in-line) |
| Temperature | $twat$ | *°C* | CU WFP (in-line) |
| $UV_{254}$ absorbance | $uv$ | *absorbance* | CU WFP (grab sample) |
| Alkalinity | $alk$ | $mg \cdot L^{-1}$ | CU WFP (grab sample) |
| Water age | $age$ | *days* | $age_t =$ days between sample collection and retrieval |

[a] All values represent daily averages unless indicated otherwise.
[b] Number of weather stations with ≥ 80% data coverage throughout the study period.
[c] Baseflow was calculated using the *baseflow* function in the R package *lfstat*.

imputation results of the time series. Missing environmental covariate data can be attributed to sporadic instrument failure and assumed to be missing at random. Missing micropollutant data can be attributed to either failure of the peak picking algorithm or presence below the limit of detection, which may or may not be missing at random. Therefore, we tested for potential violations of the missing at random assumption of multiple imputation for each MPP.[159] If a bias in the imputed results for a MPP was detected (student's t-test, $p < 0.05$), then an imputation bias correction factor (*MI_bias*) was added to the list of covariates for further examination in the multivariable regression analyses. *MI_bias* is a binary factor, where days with imputed micropollutant concentrations are assigned a value of 1, and 0 otherwise. The cumulative MPPs (*i.e.*, CUMA, CUMC, CUMC2, and CUMT) were calculated prior to multiple imputation and therefore only represent the measured results. We note that four of the 552 water samples were lost and all micropollutant data were imputed for those days.

#### *4.2.5 Data Processing – seasonal dependencies*

To account for apparent seasonal dependencies of environmental covariates on the micropollutant dynamics of agriculture-derived micropollutants, we first performed a simple slopes analysis between the environmental covariates and micropollutant concentrations using the R package *jtools*.[160] If the regression slopes were significantly different for any environmental covariate during an estimated growing season of 15 May through 15 August and outside the estimated growing season, then the growing season (*grow*) was added to the list of environmental covariates as an interaction term. Then, the apparent growing season for each micropollutant exhibiting an apparent seasonal dependence was determined by maximizing the correlation coefficient of the micropollutant concentration profile with *bfi* (a highly correlated environmental covariate) and the length of the growing season.

### *4.2.6 Data Processing – cross-correlation*

The cross-correlation coefficients of environmental covariates and MPPs were determined at different time lags by the *ccf* function in the R package *forecast*.[161] Lagged environmental covariates with greater correlation coefficients replaced the non-lagged covariates in further analyses. Environmental covariates with poor correlation ($\rho < 0.2$) or non-significant likelihood ratio tests (Chi-squared statistic, $p > 0.01$) with a MPP were removed from further analyses.

### *4.2.7 Multivariable Regression – GLS and Logit*

MPPs were evaluated by means of two independent multivariable regression techniques. First, generalized least squares (GLS) regression was conducted using the *gls* function in the R package *nlme*[162] to determine which environmental covariates are important determinants of overall micropollutant dynamics. To correct for autoregressive errors, an AR(1) correlation structure was included in the *gls* function after an examination of the (partial) autocorrelation function plots using the *acf* and *pacf* functions in the R package *forecast*.[161] Final regression parameters were examined for co-integration using the *ca.ja* function in the R package *urca*.[163] Second, logistic (logit) regression was conducted using the *glm* function with a binomial distribution model and a logit link function to determine which environmental covariates are important determinants of micropollutant peak events. Response variables for logit regression were binary transformations of the MPPs, where the greatest 25% of values measured throughout the study period were assigned a value of 1 to represent peak events, and 0 otherwise. Final regression parameters were examined graphically for linearity with the log odds of the dependent variable.

### *4.2.8 Multivariable Regression – best subsets regression.*

We used the R package *glmulti*[164] and the genetic algorithm method to find the best combinations of environmental covariates for each independent multivariable regression technique among all possible combinations and to identify important covariates for further examination. Each imputed dataset was analyzed independently. To account for multicollinearity among the environmental covariates, a variance inflation factor (VIF) filter was added to the genetic algorithm so that models with a $VIF_{max} > 3$ were excluded.[165] The models were ranked by the corrected Akaike information criterion (AICc), which is a measure of the goodness-of-fit and penalizes against overfitting.[166] For each imputed dataset, the relative importance of predictor variables were calculated using Akaike weights, which is a measure of the relative likelihood of one model being favored over another.[166] Covariates with a relative importance > 0.8 in at least three of the imputed datasets were included in downstream analyses.

### *4.2.9 Multivariable Regression – all subsets regression*

We used the R package *glmulti*[164] to conduct an all subsets regression analysis, which examines all possible combinations of important environmental covariates, for each independent multivariable regression technique and to determine statistically significant covariate coefficients. For each model within a cumulative Akaike weight of 0.99, the R package *MuMIn*[167] was used to conduct coefficient averaging and standard error adjustments using the *par.avg* function.[166] Average coefficients from each imputed dataset were then combined and their standard errors adjusted to reflect the overall variability in the model averaging and multiple imputation results.[168] The all subset regression analysis and final coefficient averaging was cross validated six times with *hv*-block cross validation.[169] The cross validation time series were selected to provide adequate coverage of the growing season in both the training and test sets.

## 4.3 Results and Discussion

### *4.3.1 Data processing*

The first step of our analysis was to examine the completeness of our overall dataset. Five of the environmental covariates (*turb*, *srad*, *pH*, *uv*, and *lwet*) contained missing data ranging between 0.2 and 2.9%. Each of the representative MPPs also contained some amount of missing data ranging between 6.2 and 63.2%. The extent of the missing values for each environmental covariate and micropollutant is provided in **Figure C3.** We used multiple imputation under the missing at random assumption to generate five complete datasets.[42,170] Although much of the missing micropollutant data was observed during periods when the actual concentration was expected to be low, the *MI_bias* terms were not found to be significant in the final regression models. This suggests that the missing at random assumption is valid. The complete temporal profiles for the representative MPPs (**Figures C4 – C11**), cumulative metrics of overall micropollutant contamination (**Figures C12 – C15**), and environmental covariates (**Figures C16 – C38**) are provided in **Appendix C**.

Some of the temporal profiles for the representative MPPs and the cumulative metrics of overall micropollutant contamination exhibited an apparent seasonal dependence. ATR, ATR.d, MET, MET.o, CUMA, CUMC, CUMC2, and CUMT exhibit significant interactions between an apparent growing season and certain environmental covariates among *flow*, *bfi*, *stp*, *pstp*, *turb*, *uv*, *alk*, *pH*, and *adp*. Therefore, the interaction term *grow* was included in the analysis for environmental covariates that exhibit significantly different relationships with the micropollutant temporal profiles during the growing and non-growing seasons. The timing of the apparent growing seasons was determined independently for each micropollutant temporal profile by maximizing the correlation coefficient with *bfi* and the length of the growing season. For example,

the apparent growing seasons for atrazine were determined to be 18 May – 4 September 2017 (109 days) and 8 May – 27 July 2018 (80 days) as shown in **Figure C39**. This approach allows us to identify approximate growing seasons for each micropollutant in the absence of actual pesticide application data.

Finally, we replaced the environmental covariates with lagged environmental covariates if the lagged environmental covariates exhibited greater correlation coefficients with the MPPs. All environmental covariates included in the final analyses were either non-lagged ($X_t$) or one-day lagged ($X_{t-1}$). Although, the differences between the lagged and non-lagged correlation coefficients were relatively small ($\Delta\rho < 0.1$).

***4.3.2 Multivariable Regression – micropollutant dynamics.***

Following data processing to generate five complete datasets, we implemented GLS multivariable regression to determine which environmental covariates are important determinants of overall micropollutant dynamics and to determine the best predictive combination of covariates. We examined multiple linear regression models that include the transformed environmental covariates for each MPP. The *MI_bias* term was added to the model equations for MPPs that exhibited potential multiple imputation bias and the *grow* interaction term was added for MPPs that exhibit an apparent seasonal dependency. Environmental covariates that exhibited poor correlation ($\rho < 0.2$) with a MPP were excluded from the respective model equation.

A best subsets regression analysis was conducted for each MPP to identify the important covariates for predicting micropollutant dynamics. For example, the model equation for implementing the GLS best subsets regression of ATR is provided as **Equation (1)**. Bivariate GLS regression plots between each MPP and covariate included in the respective best subsets regression model equations are provided in **Appendix C**. For each imputed dataset, the resulting models were

ranked by AICc to calculate relative variable importance. Some imputed datasets ranked different collinear covariates as important; therefore, we limited further analyses to covariates with greater than 0.8 relative importance in at least 3 of the imputed datasets (*e.g.*, if *tmax* was found to be important in one imputed dataset and *twat* was found to be important in four imputed datasets, *twat* would be included instead of *tmax* in further analyses). Due to the high multicollinearity among the environmental covariate data as shown in **Figure C40**, certain covariates can be interchanged with other collinear covariates; however, our method reveals which combinations of covariates can best predict the micropollutant response.

$$\begin{aligned}\log_{10}(ATR_t) = \widehat{\beta_0} + grow_t * (\widehat{\beta_1}\log_{10}(flow_t) + \widehat{\beta_2}bfi_t + \widehat{\beta_3}\log_{10}(pstp_t) \\ + \widehat{\beta_4}\log_{10}(turb_t) + \widehat{\beta_5}\log_{10}(uv_t) + \widehat{\beta_6}alk_t + \widehat{\beta_7}adp_{t-1}) \\ + \widehat{\beta_8}tavg_{t-1}{}^2 + \widehat{\beta_9}tmin_{t-1}{}^2 + \widehat{\beta_{10}}tmax_{t-1}{}^2 + \widehat{\beta_{11}}twat_{t-1}{}^2 \\ + \widehat{\beta_{12}}MI\_bias_t\end{aligned} \quad \textbf{(1)}$$

Next, an all subsets regression analysis was conducted for each MPP to determine the best predictive combination of covariates and to determine final standardized coefficient values. The model equations for the all subsets regression analysis were truncated forms of the best subsets regression equations and only included the important covariates identified during best subsets regression analysis. For example, the model equation for the implementing the GLS all subsets regression of ATR is provided as **Equation (2)**. The resulting models were again ranked by AICc and the regression coefficients were averaged. The all subsets regression analysis and model averaging was cross validated as described in **Figure C41**. Then, the final models were analyzed using the entire dataset. Further details regarding the cross validation and final coefficients are

provided in **Appendix C**. Statistically significant (t-statistic, $p < 0.05$) covariates obtained from the all subsets regression and model averaging for each MPP are reported in **Table C1** with prediction curves and $R^2$ diagnostics. The results of the GLS regression analysis showed that the environmental covariates predicted the dynamics of the representative MPPs while explaining 61 – 93% of the variance, depending on the micropollutant. Using these results, we can draw several important insights into the drivers of micropollutant dynamics in the Fall Creek watershed.

$$\log_{10}(ATR_t) = \widehat{\beta_0} + grow_t * (\widehat{\beta_1} bfi_t + \widehat{\beta_2} \log_{10}(turb_t) + \widehat{\beta_3} log_{10}(uv_t) + \widehat{\beta_4} alk_t) + \widehat{\beta_5} tmax_{t-1}{}^2 \quad \textbf{(2)}$$

Concentrations of wastewater-derived compounds such as DES, DHB, and GAB.l were mainly driven by hydrology covariates such as *flow*, *bfi*, and *pstp*. These micropollutants are likely discharged into Fall Creek through upstream STP outfalls, which is diluted during periods of high streamflow. Interestingly, *uv* was a significant predictor of DES and DHB with a negative association. In previous studies, *uv* reduction was positively associated with micropollutant removal in wastewater treatment[171] and *uv* has been suggested as a wastewater indicator in surface waters.[172] Our results suggest that any increases of *uv* caused by the presence of wastewater is likely masked by high concentrations of natural organic matter caused by surface runoff during storm events; this may be especially true for small streams with relatively low fractions of wastewater (ranged 0.6 – 2.6% at the FCMS). Another important predictor of the wastewater-derived micropollutants was *alk*, which has been shown to be a strong indicator of dilution in small streams[173,174] and was significantly correlated with *flow* during the study period ($\rho = -0.87$); to the best of our knowledge, *alk* has never before been associated with micropollutant dynamics and our

results show that *alk* can be used as a predictor of wastewater-derived micropollutant dynamics in the absence of flowrate data.

The concentration dynamics of ATR and its transformation products were mainly driven by storm events and best predicted by environmental covariates such as *bfi*, *uv*, and *turb* during the growing season. A previous study also showed that concentrations of ATR in surface water systems were negatively correlated with *bfi*.[144] This suggests that ATR, ATR.h, and ATR.d are transported to Fall Creek in overland flow during storm events. ATR exhibits a strong affinity to soils,[175] and is likely adsorbed onto soil particles that are transported into the stream, which is reflected in the positive associations with the *turb* and *uv* profiles. ATR.h has a lower tendency to leach from soils than ATR and ATR.d,[176] which likely explains the lack of a significant *grow* interaction term and the concentration dynamics of ATR.h extending beyond the ATR growing season. Similarly, *bfi* and *uv* were found to be good predictors of MET and MET.o concentration dynamics during the growing season. Our analysis revealed that *uv* was the most important and most consistent predictor of agriculture-derived micropollutants and therefore a good indicator of pesticide concentration dynamics during the growing season in surface waters. Generally, the agriculture-derived micropollutants were present at higher concentrations during the growing season, which is reflected in the high positive coefficient values of *grow* and temperature covariates (*tmax* and *twat*). For all agricultural MPPs, the significant interaction terms are higher during the growing season and lower or have opposite signs during the non-growing season. For example, *turb* is positively associated with ATR during the growing season and negatively associated during the non-growing season. Smaller associations with environmental covariates during the non-growing season suggest that the micropollutant is persistent in soil, which is

transported to Fall Creek during storm events; opposite associations suggest pesticide persistence in groundwater recharge, which is diluted during storm events.

The GLS regression performed poorly for the cumulative micropollutant metrics, which represent combinations of different MPPs with many different sources and drivers of their mobilization and transport. The GLS regression is applicable for individual micropollutants because the emissions of the representative micropollutants are well-defined and each have a single set of fate and transport processes that govern their concentration dynamics. Our results show that the concentration dynamics of particular micropollutant classes (wastewater and agriculture-derived) have similar predictive environmental covariates that likely control the fate, transport, and concentrations of those micropollutant classes. However, because the cumulative micropollutant metrics contain different micropollutant classes, the GLS regression analysis fails to identify predictive covariates because the underlying fate and transport processes for each micropollutant are commingled within each cumulative profile. Because the modeling framework for the micropollutant dynamic GLS regression did not perform well for predicting the dynamics of the cumulative MPPs, we investigated if the environmental covariates could be used to predict peak events associated with MPPs and the cumulative micropollutant metrics.

### *4.3.3 Multivariable Regression – peak events*

We implemented logit multivariable regression to determine which environmental covariates are important determinants of micropollutant peak events. We define peak events as the days when the value of the temporal profile is in the upper quartile (*i.e.*, top 25%) of all measured values. The analysis was limited to the growing season for MPPs where all peak events occurred during the growing season (ATR, ATR.d, and MET), and peak events were determined separately for the growing and non-growing seasons for MPPs where peak events occurred during the

growing and non-growing seasons (CUMC, CUMC2, and CUMT). Similar to the GLS regression analysis, multiple linear regression models were constructed for each MPP. Environmental covariates that exhibited nonsignificant likelihood ratio tests, which compare bivariate models against the null model, were excluded from the respective logit model equation.

A best subsets regression analysis was conducted for each MPP to identify the important covariates that predict peak events. For example, the model equation for implementing the logit best subsets regression of ATR is provided as **Equation (3)**. Bivariate logit regression plots for each MPP and covariate included in the respective best subsets regression model equations are provided in **Appendix C**.

$$\begin{aligned} Pr(ATR_t) = \widehat{\beta_0} + grow_t * (\widehat{\beta_1}\log_{10}(flow_t) + \widehat{\beta_2}bfi_t + \widehat{\beta_3}\log_{10}(pstp_t) \\ + \widehat{\beta_4}\log_{10}(turb_t) + \widehat{\beta_5}\log_{10}(uv_t) + \widehat{\beta_6}alk_t + \widehat{\beta_7}pH_t + \widehat{\beta_8}adp_{t-1} \\ + \widehat{\beta_9}tavg_t{}^2 + \widehat{\beta_{10}}tmin_t{}^2 + \widehat{\beta_{11}}tmax_t{}^2 + \widehat{\beta_{12}}twat_t{}^2 + \widehat{\beta_{13}}lwet_t \\ + \widehat{\beta_{14}}rhum_{t-1} + \widehat{\beta_{15}}srad_t + \widehat{\beta_{16}}prcp01_t) \end{aligned} \quad \textbf{(3)}$$

Where $Pr(ATR_t) = \ln(\frac{P_t}{1-P_t})$ and $P_t$ is the probability of a peak event on a given day.

The same procedure as the GLS regression analysis was followed to determine which covariates are considered important predictors of peak events for each MPP and to construct model equations for an all subsets regression analysis. The model equation for implementing the logit all subsets regression of ATR is provide as **Equation (4)**. The resulting models were averaged and statistically significant (z-statistic, $p < 0.05$) covariates for each representative and cumulative MPP are reported in **Tables C1** and **C2**, respectively, with receiving operating characteristic (ROC) curves and area under curve (AUC) diagnostics. The results of the logit regression analysis

showed that the environmental covariates accurately predicted the peak events of the representative MPPs (0.83 – 0.94 AUC) and cumulative MPPs (0.82 – 0.92 AUC). The logit regression analysis for the representative MPPs identified similar predictive environmental covariates as the GLS regression analysis, which suggests that this method can also be used to improve our fundamental understanding of the drivers of micropollutant dynamics. Additionally, the logit regression analysis performed well for the cumulative MPPs, which allows us to draw important insights into the drivers of overall micropollutant contamination in surface waters.

$$Pr(ATR_t) = \widehat{\beta_0} + grow_t * (\widehat{\beta_1} bfi_t + \widehat{\beta_2} alk_t + \widehat{\beta_3} adp_{t-1} + \widehat{\beta_4} tmax_t{}^2) \quad \textbf{(4)}$$

Where $Pr(ATR_t) = \ln(\frac{P_t}{1-P_t})$ and $P_t$ is the probability of a peak event on a given day.

The dominant predictor of peak events for wastewater-derived micropollutants was *flow*. Other significant predictors, *alk*, *uv*, and *adp*, are associated with *flow* and dilution changes in Fall Creek. These results show that the highest concentrations of wastewater-derived micropollutants occur during periods of low streamflow, which has been demonstrated in previous studies.[35,126,177] However, using our highly resolved MPPs we are able to use specific streamflow-related covariates to accurately predict when peak events would occur.

Peak agriculture-derived micropollutant concentrations occurred more often during the growing season as indicated by the positive associations with *grow*, *srad*, and temperature covariates; the dominant predictor was *bfi* during the growing season. These results show that the highest concentrations of agriculture-derived micropollutants occur during periods of high streamflow during the growing season, also agreeing with expectations and previous studies.[35,126]

Nevertheless, our results indicate that *bfi*, along with *adp*, *alk*, *turb*, and *uv*, can most accurately predict peak events of agriculture-derived micropollutants.

The peak events for cumulative MPPs were well predicted by the set of environmental covariates. It is important to note that the cumulative micropollutant metrics are biased towards the specific micropollutants that were previously prioritized, quantified, or had available toxicity data. Nevertheless, these types of overall micropollutant contamination metrics have been examined previously,[13,32,98] but never before at high temporal resolution. Peak events for cumulative MPPs were dominated by agriculture-derived micropollutants during the growing season and wastewater-derived micropollutants at other times. The cumulative MPPs contain other micropollutant classes with less defined emissions, but our results suggest that their peak events are either less intense or driven by similar processes as the representative micropollutant classes. Positive associations with *turb* (for CUMA and CUMC2) and negative associations with *bfi* (for CUMC and CUMT) best predicted the peak events during the growing season and associations with dilution covariates (*alk*, *flow*, and *bfi*) best predicted the peak events during the non-growing season, which is similar to the results of the representative MPPs. The risk associated with high concentrations of micropollutants in Fall Creek is driven by runoff events during the growing season and changes in dilution throughout the year. Generally, peak events occurred more frequently and at higher levels of contamination during the growing season due to the addition of many transient pesticides in the Fall Creek watershed.

The cumulative risk profile is best predicted by negative associations with *bfi* and *alk* during the growing season and positive associations with *alk* during the non-growing season. We calculated a maximum cumulative EAR value of approximately 0.05. Previous studies have shown wide ranges of cumulative EAR values depending on sampling location and detected

micropollutants (ranging from approximately $10^{-3}$ to 60); a maximum cumulative EAR of 0.03 has been reported in drinking water resources.[152] It is worth noting that EAR-based assessments of ToxCast-specific chemical risk may underestimate the potential for cumulative effects by at least 2-3 orders of magnitude.[13] Therefore, cumulative EAR values of 0.001 and 0.01 have been suggested as thresholds to minimize the potential for toxic effects within sensitive populations.[13] EAR values exceeded the 0.01 thresholds in 14% of our samples, suggesting that cumulative toxicity might be of concern during peak events in Fall Creek. Our modelling approach has allowed us to identify environmental covariates that reasonably predict peak events and the results may be used to inform more accurate chemical risk assessments in Fall Creek and similar surface water systems.

#### *4.3.4 Recommendations for future micropollutant sampling strategies*

We next investigated whether our models could be used to develop intermittent sampling strategies that provide a more accurate representation of chemical risk. Micropollutant monitoring in surface waters is becoming more widespread around the world;[3,32,35,98,178–180] however, highly resolved temporal sampling to assess micropollutant dynamics is still rarely implemented.[34,146] On the contrary, environmental covariate data including weather, hydrology, and water quality parameters are routinely measured and recorded, and are often made freely available. Here, we explore our complete dataset to evaluate whether certain intermittent sampling strategies could be implemented to more accurately represent chemical risk associated with micropollutants.

We first examined whether an informed sampling strategy that targets peak events is more likely to represent the true extent of micropollutant dynamics and environmental risk when compared to an uninformed sampling strategy. To do this, we first divided each of our MPPs into weekly and monthly intervals. We then drew samples from each of the weekly or monthly intervals

in either an informed or an uninformed way. We defined an informed sampling strategy by using the results of the logit regression analysis and the important bivariate models to identify samples within each interval that exhibited the highest probability of a peak event. The uninformed sampling strategy was implemented by drawing a sample at a fixed interval (*i.e.*, every 7 or 30 days starting on each day within the first interval) regardless of the prevailing environmental conditions, which is a commonly implemented micropollutant sampling strategy.[44,87,96,98,181] As a means to benchmark the accuracy of the informed and uninformed sampling strategies, we also identified the maximum value for each MPP in each sampling interval. We finally compared the performance of the informed and uninformed sampling strategies by calculating the percentage of the actual environmental risk captured by each sampling strategy. The results of this exercise are described in **Table C3** for ATR, DES, CUMC, CUMC2, and CUMT. **Figures 4.2** and **C62 – C65** show the resulting MPPs obtained from each sampling strategy. Our results indicate that informed sampling can significantly improve the accuracy of environmental risk assessments; the percent of environmental risk captured using the informed sampling strategies is greater than the average risk captured by the uninformed strategies. The improvement is magnified for more dynamic MPPs (*e.g.*, ATR, CUMC2, CUMT) and for longer sampling intervals (*i.e.*, monthly sampling versus weekly sampling) as exemplified in **Figure 4.2**. This informed sampling strategy captured 106% and 99% of the actual maximum atrazine concentrations during monthly and weekly sampling, respectively, versus an average of only 29% and 53% for the uninformed sampling strategy (see **Table C3**). Based on this analysis, we recommend informing micropollutant sampling strategies based on measurements of *bfi*, *alk*, *flow*, or *uv*.

As another approach to informing micropollutant sampling strategies to target peak events, we identified values of specific environmental covariates that could be used as triggers for surface

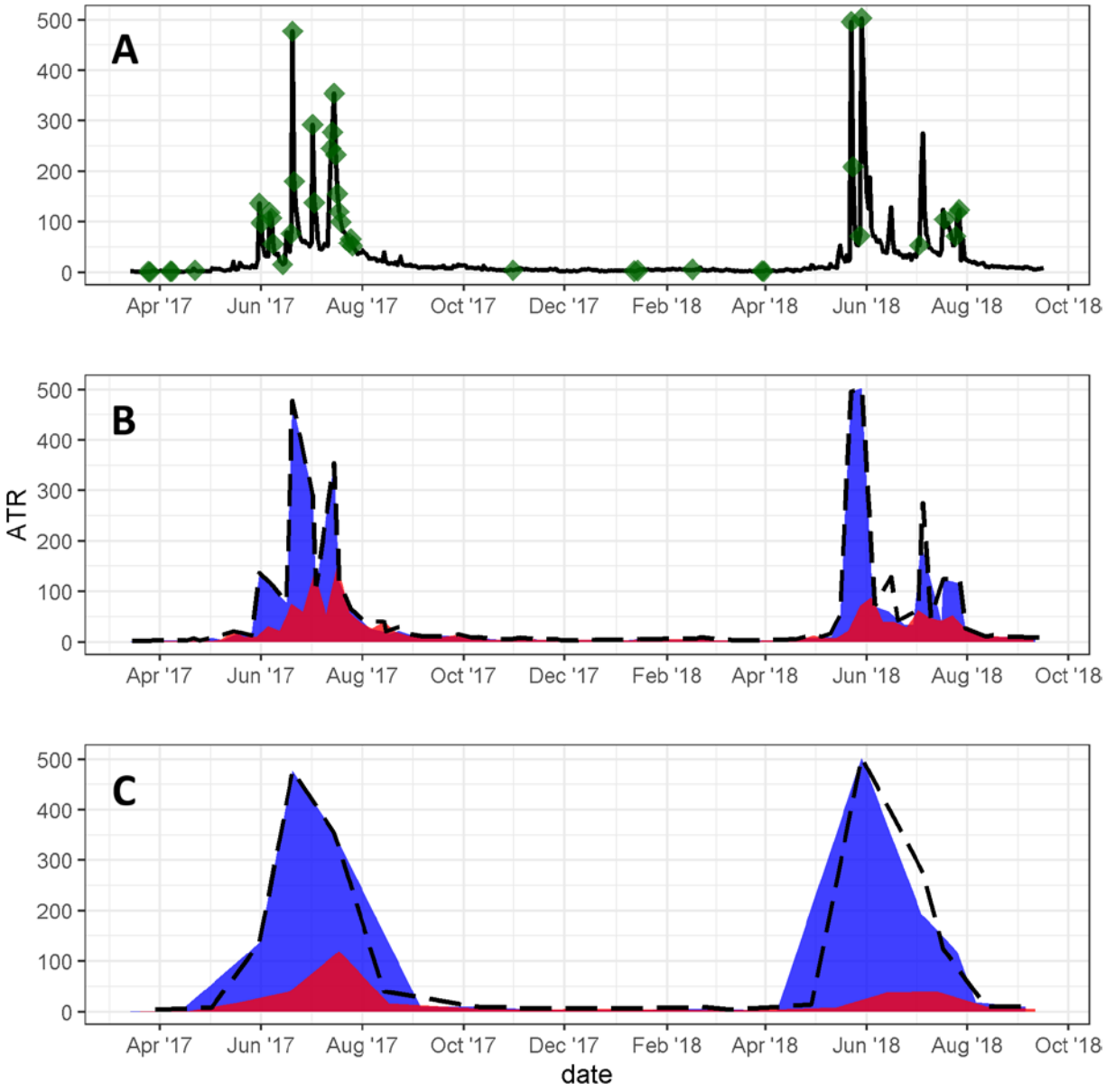



**Figure 4.2:** Informed vs uninformed sampling strategies for atrazine (ATR) concentration in $ng \cdot L^{-1}$: A) daily, B) weekly, and C) monthly; solid black line = average of the imputed datasets; green points = triggered sampling events (≥2); dashed black line = actual representation of risk; blue area = informed sampling, red area = uninformed sampling.

water sampling. We used the important bivariate models to identify the values of each environmental covariate at which the probability of a peak event exceeds 50%. Values are reported as absolute values for Fall Creek and as percentiles of the overall dataset as a means to generalize our results to other similar surface water systems. These values are provided in **Table 4.2**, and the green diamonds in **Figures 4.2** and **C62 – C65** identify the days at which at least two of the environmental covariates exceed the threshold values. The number of environmental covariate threshold exceedances for two simultaneous triggers ranged from 36 – 90 days depending on the MPP and correctly selected a peak event day 55 – 73% of the time, which is an improvement over uninformed sampling strategies that randomly select days to sample (25% of all days are defined as peak events). We contend that these data can be used as triggers for micropollutant sampling in similar watersheds. For example, we have shown that the probability of a peak event occurring for agriculture-derived micropollutants during the growing season exceeds 50% when *turb* is greater than 20 NTU in Fall Creek, which is the 87$^{th}$ percentile value for *turb* in Fall Creek; we expect that the percentile value will be more valuable than the absolute value when extending our results to other similar surface water systems.

Finally, we aimed to use our results to contextualize micropollutant data derived from intermittent grab sampling. A previous study conducted by others reported concentrations of atrazine (381 ng·L$^{-1}$) and desvenlafaxine (< 7.4 ng·L$^{-1}$) in a grab sample collected from Fall Creek during the agricultural growing season in 2014.[32] The streamflow rate at the reported sampling time was 21 m$^3$·s$^{-1}$, which exceeds the threshold for a predicted peak event for agriculture-derived micropollutants during the growing season but not for wastewater-derived micropollutants. Therefore, the reported micropollutant results can be contextualized as representing a relatively high concentration of atrazine and a relatively low concentration of desvenlafaxine. The reported

concentrations also agree in magnitude with values reported for peak atrazine and non-peak desvenlafaxine in the present work. This analysis provides an example of how the data in **Table 4.2** can be used to help interpret micropollutant data derived from intermittent grab samples.

**Table 4.2:** Environmental covariate triggers for micropollutant sampling aimed at capturing peak events in surface waters. Measured values are reported for Fall Creek and as percentiles of the overall dataset.

| Micropollutant Profile | | BFI (-) | Flowrate ($m^3 \cdot s^{-1}$) | Turbidity (NTU) | $UV_{254}$ (absorbance) | Alkalinity ($mg^3 \cdot L^{-1}$) | Precipitation (mm) |
|---|---|---|---|---|---|---|---|
| **Agricultural-derived** | *Growing season* | < 0.40 (30%) | > 7.3 (77%) | > 20 (87%) | > 0.18 (89%) | < 120 (35%) | > 18[a] (96%) |
| | *Non-growing season* | < 0.04 (<1%) | > 32 (99%) | > 64 (97%) | > 0.30 (97%) | < 68 (<1%) | > 25[a] (98%) |
| **Wastewater-derived** | -- | > 0.93 (86%)[a] | < 1.6 (22%)[a] | < 2.3 (5%) | < 0.05 (<1%) | > 156 (85%) | *n.a.* |
| **Cumulative concentration (CUMC)** | *Growing season* | < 0.30 (20%)[a] | *n.a.* | > 40 (94%) | > 0.24 (95%) | *n.a.* | > 16[a] (95%) |
| | *Non-growing season* | > 0.96 (89%)[a] | < 1.4 (19%)[a] | *n.a.* | *n.a.* | > 156 (85%) | *n.a.* |
| **Cumulative concentration (CUMC2)** | *Growing season* | < 0.38 (29%) | > 6.7 (75%) | > 17 (85%) | > 0.16 (86%) | < 119 (34%) | > 13[a] (94%) |
| | *Non-growing season* | > 0.84 (78%)[a] | < 1.0 (7%)[a] | < 1.5 (<1%) | < 0.04 (<1%) | > 169 (98%) | *n.a.* |
| **Cumulative toxicity (CUMT)** | *Growing season* | < 0.47 (39%) | > 6.2 (73%) | > 16 (84%) | > 0.14 (80%) | < 119 (34%) | > 14[a] (95%) |
| | *Non-growing season* | *n.a.* | < 0.86 (3%)[a] | *n.a.* | > 0.8 (99%) | > 166 (96%) | *n.a.* |

[a] one-day lagged

n.a. = not applicable; unimportant covariate for the respective micropollutant profile.

# CHAPTER 5 – Conclusions

## 5.1 Environmental micropollutant monitoring

The two micropollutant monitoring campaigns described in this dissertation were designed to comprehensively characterize micropollutant contamination in a way that can lead to best management practices (BMPs) for micropollutant management at a watershed scale. Although each campaign was designed to address specific research objectives, both designs have certain advantages and disadvantages. The two campaigns differ in sampling and micropollutant characterization design. The design considerations for the Hudson River Estuary (HRE) and Fall Creek (FC) micropollutant monitoring campaigns are summarized in **Table 5.1**. Design considerations can limit the scope of the resultant datasets and must be selected carefully to address the main research objectives.

**Table 5.1:** A comparison of the design considerations of the two micropollutant monitoring campaigns.

| Design consideration | Hudson River Estuary (Project A) | Fall Creek (Projects B and C) |
|---|---|---|
| Surface water system | Large industrialized estuary and mixed-use tributaries | Small rural stream in a mixed-use watershed |
| Sampling strategy | Spatially distributed (17 sites & 9 sampling dates) | Temporally distributed (1 site & 552 sampling dates) |
| Sampling method | Grab (manual aboard boat) | Time-proportional composite (autosampler) |
| Sample preparation | Filter and pH adjustment | Centrifuge and pH adjustment |
| Sample enrichment | Solid phase extraction, x1000 (manual) | Large volume injection, x250 (automatic) |
| Micropollutant characterization type | Target screening by HPLC-HRMS/MS | Target and nontarget screening by HPLC-HRMS(/MS) |
| Micropollutant data processing method | XCalibur (vendor software) | enviMass and custom R-scripts (open source software) |
| Environmental covariates | Geospatial features and streamflow rates | Weather, hydrology, and water quality parameters |

### *5.1.1 Sampling design*

The most critical sampling design consideration is the sampling strategy – where and when to sample, which defines the scope of the resultant datasets from the outset of analysis. Spatiotemporal sample resolution should be defined by the research objective. For example, spatially distributed sampling strategies may enable source attribution[36,98] and temporally distributed strategies can capture micropollutant dynamics.[37,146] The monitoring campaigns employed in this dissertation were designed to collect datasets that can be used to make recommendations for micropollutant management at the watershed scale. Using a spatially distributed sampling strategy, data from the HRE campaign allowed me to identify specific tributaries of concern that contribute high loads of different types of micropollutants. Using a temporally distributed sampling strategy, data from the FC campaign allowed me to provide specific recommendations on when peak contamination events are likely to occur based environmental covariate data.

Sampling methods can alter the perceived micropollutant contamination at a particular site. Grab sampling is the easiest sampling method, especially for highly resolved spatial sampling, but may misrepresent the true micropollutant contamination; grab samples can only provide a snapshot of variable micropollutant dynamics.[38,182] However, data derived from a series of grab samples can provide robust estimates of the true micropollutant contamination at a particular sample site. Flow or time-proportional composite sampling can more accurately reflect micropollutant contamination, but require an autosampler and access to power and shelter.[38,182] The HRE campaign utilized grab sampling due to the widespread spatial component of the sampling strategy that extended over 200 km of river and hard-to-reach sample sites. The FC campaign utilized an autosampler to automatically collect daily time-proportional samples, which required minimal

manual interaction throughout the 18 month sampling campaign and permitted the longevity of the study period.

Sample enrichment methods are critical because many micropollutants are present at trace concentrations in aquatic systems, which may be below analytical limits of detection.[183,184] For multiresidue analyses, it is important that the sample enrichment method can capture a broad range of micropollutants.[185] The HRE campaign utilized solid phase extraction (SPE) to achieve high concentration factors and simultaneously provide matrix cleanup of relatively polluted water samples.[81,186] The FC campaign utilized large volume injection (LVI) due to the time constraints associated with the large number of relatively pristine water samples analyzed during the study.[118] In these cases, the SPE method is more time-intensive and costly, but achieves a higher concentration factor than the LVI method. The decisions regarding sampling design will affect the micropollutant characterization results, which can only measure what is present in the final sample.

### *5.1.2 Micropollutant characterization design*

The micropollutant characterization design considerations, including the type of micropollutant screening, micropollutant data processing method, and the inclusion of environmental covariates, will define the scope of resultant datasets. The micropollutant characterization methods described in this dissertation utilized high performance liquid chromatography (HPLC) coupled with high-resolution mass spectrometry (HRMS) to conduct target and nontarget screenings. The analytical methods employed were designed to capture a wide breadth of potential water contaminants; however, the scope is limited to those micropollutants that are detectable (polar to semi-polar micropollutants that are ionizable under electrospray ionization).[81,82] Target analyses can only measure the micropollutants included in the target list and require pure analytical reference standards. Nontarget analyses (NTAs) are more complicated

and require additional data processing steps, but can more broadly characterize micropollutant occurrence and do not necessarily require pure reference standards; although, the use of reference standards is required to unequivocally confirm detections and to conduct quantitative analyses.[34,61] Furthermore, target screenings can take advantage of relatively easy-to-use vendor software, though proprietary algorithms for nontarget peak picking and prioritization do not afford the transparency needed to fully evaluate nontarget MS features. Open source software allows for the flexibility and customizability needed for data-driven NTA workflows. The HRE campaign utilized a target screening to quantify the spatiotemporal occurrence and concentration patterns of 200 micropollutants that are known or suspected to occur in surface water systems using a vendor software. The FC campaign utilized a target and nontarget screening approach to simultaneously detect target micropollutants while acquiring data for a novel NTA, which was conducted using open source R packages and custom R scripts. Lastly, by collecting environmental covariates along with micropollutant characterization data, advanced data-driven methods can be employed to further our understanding of micropollutant sources, fate, and transport. Spatial patterns of micropollutants linked to geospatial features of the watershed led to source attribution in the HRE. Temporal patterns of micropollutants linked to environmental covariates led to insights about the drivers of micropollutant dynamics in FC.

### 5.2 Outlook

The research described in this dissertation improves our environmental micropollutant monitoring capabilities to more comprehensively characterize micropollutant contamination and furthers our understanding of the sources, fate, and transport of micropollutants in surface waters by coupling broad micropollutant monitoring afforded by HRMS with data-driven methods. For

example, in Project A, I identified similar clusters of co-occurring micropollutants based on widespread micropollutant occurrence and concentration data, and linked those clusters with geospatial features of the watershed to identify sources of micropollutants. In Project B, I collected a novel micropollutant monitoring dataset in a small stream, designed a NTA workflow to tentatively identify unknown micropollutants, and revealed the temporal dynamics of micropollutant occurrence in a small stream at high resolution for the first time. In Project C, I discovered insights on the drivers of micropollutant dynamics in surface waters, and provided generalizable recommendations to water quality stakeholders on when peak micropollutant events are likely to occur based on easily obtained environmental covariates.

These project outcomes can be applied directly to the investigated watersheds and are likely conserved across similar surface water systems. For example, the compositions of the micropollutant clusters reported in Project A were shown to be conserved across global surface water systems through an extensive literature review. Therefore, the micropollutant clusters can be used as a means to select one or a few micropollutants to monitor for evidence of impact from a particular source in surface water systems around the world. Similarly, the distinct concentration-streamflow relationships discovered for agriculture and wastewater-derived micropollutants in Projects B and C were also shown to be conserved across similar surface water systems. Although specific relationships, such as the overall micropollutant contamination metrics discussed in Project A and the concentration-environmental covariate relationships discussed in Project C, may need to be recalibrated before being directly applied in other surface water systems. Additionally, the relationships discovered in Projects B and C are applicable to small streams in mixed-use watersheds with agriculture activities and/or STP discharges, but future research is required to investigate if the relationships are masked or equalized in larger surface water systems.

I exposed knowledge gaps pertaining to environmental micropollutant monitoring. First, relatively few studies have utilized data-driven methods to link geospatial data with micropollutant characterization data. I demonstrated that coupling broad micropollutant monitoring afforded by HRMS with data-driven methods is capable of furthering our understanding of the sources, fate, and transport of micropollutants in surface waters and should continue to be utilized in future studies. Second, NTAs are increasingly utilized for environmental micropollutant monitoring, but the widespread usage of NTAs is still hindered by the lack of high throughput workflows. I developed a novel NTA method and prioritization technique; however since our analytical capabilities are constantly advancing, the workflow will likely be improved upon in the near future. Third, much of the recent environmental micropollutant monitoring research is aimed at improving our knowledge of environmental micropollutant contamination, but is based on datasets that may not be representative of the true micropollutant contamination, as discussed in the previous section. I offer several ways to generalize or contextualize micropollutant characterization results, and make recommendations for future environmental micropollutant monitoring to obtain more representative datasets.

Future research should focus on obtaining results that can be used to directly inform risk assessments and eventual water quality regulations so that we can continue to protect human and ecological health against the adverse effects of micropollutant exposure. Past and current research has focused on obtaining micropollutant characterization datasets for a variety of environmental systems. However, as a research community, we need to examine the relevance of such datasets. The state-of-the-science is rapidly expanding with respect to NTA; 77% of the 490 peer-reviewed research articles regarding NTA were published within the past five years[d] and numerous

[d] Web of Science search for "ALL = (nontarget OR non-target) AND ALL = (HRMS OR high-resolution mass spectrometry)" on 8 April 2019.

micropollutants have been identified in the environment for the first time.[41,54–57,146] However, we do not yet have methods capable of quickly and effectively translating this wealth of new micropollutant monitoring data into estimates of environmental risk. New research focused on improving micropollutant monitoring approaches to broadly characterize micropollutant occurrence should also focus on informing environmental risk assessments. For example, effect-directed analysis (EDA) is a method to identify micropollutants causing adverse effects by coupling HRMS with bioassays (*i.e.*, testing the biological activity of a sample).[182] By utilizing data-driven methods to link geospatial data with micropollutant characterization data and EDA, we can not only further our knowledge about micropollutant sources, fate, and transport, but also our knowledge about the sources and drivers of environmental risk associated with micropollutants. Instead of measuring the toxic effects associated with micropollutant exposure using *in vivo* or *in vitro* methods to inform environmental risk assessments, cheminformatics tools could be used to predict toxicity using *in silico* methods.[187] Modeling using cheminformatics methods to predict the adverse effects of micropollutants on human and ecological health are based on chemical descriptors that describe physiochemical and molecular properties of micropollutants. Data-driven cheminformatics tools capable of performing environmental risk assessments could provide relevance to environmental micropollutant monitoring datasets. The research presented in this dissertation culminated in the development of a highly resolved temporal profile for estimated risk in a small stream; however, despite using the best available methods and up-to-date databases, our results were biased towards those micropollutants that are well studied due to the lack of toxicological data for most of the detected micropollutants and the lack of toxicological predictive tools. High throughput micropollutant monitoring and risk assessments using data-driven EDA

and/or cheminformatics methods could help focus micropollutant mitigation strategies and regulations on the most toxic micropollutants and their sources.

Several challenges still exist for the establishment of routine high throughput environmental micropollutant monitoring, which is needed to most accurately inform risk assessments and water quality regulations. Population growth has amplified global chemical manufacturing and production, and novel micropollutants are entering our water resources.[11] Proactively monitoring for micropollutants in the environment and identifying micropollutants of concern before symptoms are shown in the environment or in people could help prevent water quality crises such as fish population collapse after exposure to synthetic estrogen[20] and human cancer cases after exposure to PFASs.[22] Major knowledge gaps currently preventing the implementation of high throughput environmental monitoring that is capable of environmental risk assessments include the lack of automated methods to decrease time-intensive sample preparation and analysis methods, and the lack of universal open-source databases to support micropollutant monitoring. The automation of analytical and post-acquisition data processing methods would allow for water samples to be quickly prepared and analyzed, and for micropollutants with high environmental risk to be detected and confirmed. Fully automated methods from sample collection to nontarget micropollutant confirmation would allow for near continuous monitoring of micropollutants in facilities such as wastewater or drinking water treatment plants. Automated systems capable of identifying nontarget micropollutants at high spatiotemporal resolution could more easily identify concentration trends of novel micropollutant occurrence. There is an increasing need for open-access mass spectral databases to support the identification of unknown micropollutants and for open-access toxicological databases to support environmental risk assessments associated with micropollutant exposure. Several mass spectral databases are

currently available, as discussed in Project B; however, there is a need to unify these databases in order to make them more easily assessable to end-users and help optimize future NTA workflows. The Environmental Protection Agency (EPA) maintains a chemical database including results from numerous toxicity assays, as discussed in Project C; however, there are too many micropollutants in existence to comprehensively collect toxicological data for each micropollutant. Using EDA based on new monitoring data and cheminformatics methods based on available toxicological data will result in more comprehensive risk assessments. Continuous micropollutant monitoring is needed to detect high contamination events associated with the highest environmental risk and to inform the public about water quality concerns.

I believe that we must make robust scientific decisions regarding micropollutant regulation and mitigation strategies based on the limits of human knowledge. The global micropollutant situation is gaining more attention from water quality stakeholders due to the increasing amounts of research over the past few decades aimed at improving our knowledge of environmental micropollutant occurrence. In the absence of continuous micropollutant monitoring, we must use our current knowledge and analytical technology to best protect the environment and human health. Our knowledge of micropollutants in the environment is constantly evolving and the decisions we make regarding regulations and BMPs should reflect this new knowledge.

Micropollutant regulation should combine the United States and European Union (EU) regulatory approaches by regulating known micropollutants based on available occurrence and toxicology data, and regulating novel or unexpected micropollutants with unknown effects based on the precautionary principal. We should continue to perform toxicity assays for known micropollutants and estimate the toxicity associated with unexpected micropollutant exposure using cheminformatics tools to set maximum allowable concentration values. The surrogate zero

concentration value of 0.1 ug·$L^{-1}$ for individual pesticide concentrations in drinking water was set in the EU in 1980 and retained in 1998; however, modern analytical technologies can detect micropollutants in the sub-ng·$L^{-1}$ range. We need to reassess the surrogate zero concentration value for micropollutants without toxicological data and base it on current analytical technologies and the broader scientific understanding of micropollutant exposure.

We should also reassess how micropollutants are regulated with respect to which specific micropollutants are regulated and when sampling for regulatory testing should occur. First, regulations are often based on indictor micropollutants that are expected to be representative of overall contamination. For example, trihalomethanes (THMs) and haloacetic acids (HAAs) are used as indicators for disinfection byproducts, and perfluorooctanoic acid (PFOA) and perfluorooctanesulfonic acid (PFOS) are used as indicators for PFASs. Results from this dissertation and other comprehensive micropollutant monitoring studies can be used to support the selection of indicator micropollutants for broad micropollutant contamination from various sources. Overall micropollutant contamination indicators that are capable of describing cumulative concentrations of micropollutants[32,98] could be used to regulate broad micropollutant occurrence in impacted drinking water sources, as described in Project A. Source-specific indicators could be used to regulate specific groups of micropollutants such as agricultural and wastewater-derived micropollutants,[98,146] as described in Projects A and B. Future comprehensive micropollutant monitoring campaigns will add additional support for the selection of indicator micropollutants that are representative of overall micropollutant contamination in widespread surface water systems. Second, infrequent uninformed grab sampling can drastically misrepresent the true micropollutant contamination. Regulation testing should aim to capture peak contamination events

that represent maximum micropollutant concentrations associated with the highest environmental risk; Project C offers several recommendations on how to inform intermittent sampling strategies.

Furthermore, BMPs for micropollutant management at the watershed scale should be implemented at micropollutant emission sources to protect the environment and at drinking water recourses to further protect human health. Cost-effective and sustainable micropollutant mitigation strategies and treatment technologies should be employed wherever feasible at wastewater treatment plants (*e.g.*, constructed wetlands,[188] groundwater recharge[189]), diffuse emissions sources (engineered hyporheic zones[190]), and in drinking water treatment plants (biofiltration,[118] advanced adsorption technologies[191]). More advanced treatment technologies (*e.g.*, advanced oxidation processes,[192] membrane filtration[193]) capable of high removal efficiencies of broad types of micropollutants should be principally implemented in areas of high emissions or in vulnerable environments. Based on the results from Project A, Rondout Creek should be prioritized for implementation of micropollutant mitigation strategies at both wastewater treatment plants and sources of diffuse agriculture-derived micropollutant emissions. Based on the results from Projects B and C, implementation of micropollutant mitigations strategies in the FC watershed should be focused on controlling diffuse emissions of agricultural micropollutants.

Throughout this dissertation, I have shown that our environmental micropollutant monitoring capabilities can be improved by coupling broad micropollutant characterization afforded by HRMS with data-driven methods. The research outcomes directly inform water quality stakeholders for micropollutant management at the watershed scale and offer ways to generalize the results for global aquatic systems. The environmental monitoring techniques developed here and their results will aid future micropollutant monitoring campaigns obtain more representative results and enable better management of micropollutants in surface water systems.

# REFERENCES


(1) Schwarzenbach, R. P.; Escher, B. I.; Fenner, K.; Hofstetter, T. B.; Johnson, C. A.; Gunten, U. Von; Wehrli, B. The Challenge of Micropollutants in Aquatic Systems. *Science (80-. ).* **2006**, *313* (5790), 1072–1077.

(2) Richardson, S. D.; Ternes, T. A. Water Analysis: Emerging Contaminants and Current Issues. *Anal. Chem.* **2018**, *90* (1), 398–428. https://doi.org/10.1021/acs.analchem.7b04577.

(3) Kolpin, D. W.; Furlong, E. T.; Meyer, M. T.; Thurman, E. M.; Zaugg, S. D.; Barber, L. B.; Buxton, H. T. Pharmaceuticals, Hormones, and Other Organic Wastewater Contaminants in U.S. Streams, 1999-2000: A National Reconnaissance. *Environ. Sci. Technol.* **2002**, *36* (6), 1202–1211. https://doi.org/10.1021/es011055j.

(4) Phillips, P. J.; Schubert, C.; Argue, D.; Fisher, I.; Furlong, E. T.; Foreman, W.; Gray, J.; Chalmers, A. Concentrations of Hormones, Pharmaceuticals and Other Micropollutants in Groundwater Affected by Septic Systems in New England and New York. *Sci. Total Environ.* **2015**, *512–513*, 43–54. https://doi.org/10.1016/j.scitotenv.2014.12.067.

(5) Fenner, K.; Canonica, S.; Wackett, L. P.; Elsner, M. Evaluating Pesticide Degradation in the Environment: Blind Spots and Emerging Opportunities. *Science (80-. ).* **2013**, *341* (6147), 752–758. https://doi.org/10.1126/science.1236281.

(6) Han, E. J.; Lee, D. S. Significance of Metabolites in the Environmental Risk Assessment of Pharmaceuticals Consumed by Human. *Sci. Total Environ.* **2017**, *592*, 600–607. https://doi.org/10.1016/j.scitotenv.2017.03.044.

(7) Evgenidou, E. N.; Konstantinou, I. K.; Lambropoulou, D. A. Occurrence and Removal of Transformation Products of PPCPs and Illicit Drugs in Wastewaters: A Review. *Sci. Total Environ.* **2015**, *505*, 905–926. https://doi.org/10.1016/j.scitotenv.2014.10.021.

(8) Benner, J.; Helbling, D. E.; Kohler, H. P. E.; Wittebol, J.; Kaiser, E.; Prasse, C.; Ternes, T. A.; Albers, C. N.; Aamand, J.; Horemans, B.; Springael, D.; Walravens, E.; Boon, N. Is Biological Treatment a Viable Alternative for Micropollutant Removal in Drinking Water Treatment Processes? *Water Res.* **2013**, *47* (16), 5955–5976. https://doi.org/10.1016/j.watres.2013.07.015.

(9) Luo, Y.; Guo, W.; Ngo, H. H.; Nghiem, L. D.; Hai, F. I.; Zhang, J.; Liang, S.; Wang, X. C. A Review on the Occurrence of Micropollutants in the Aquatic Environment and Their Fate and Removal during Wastewater Treatment. *Sci. Total Environ.* **2014**, *473–474*, 619–641. https://doi.org/10.1016/j.scitotenv.2013.12.065.

(10) Glassmeyer, S. T.; Furlong, E. T.; Kolpin, D. W.; Batt, A. L.; Benson, R.; Boone, J. S.; Conerly, O.; Donohue, M. J.; King, D. N.; Kostich, M. S.; Mash, H. E.; Pfaller, S. L.; Schenck, K. M.; Simmons, J. E.; Varughese, E. A.; Vesper, S. J.; Villegas, E. N.; Wilson, V. S. Nationwide Reconnaissance of Contaminants of Emerging Concern in Source and Treated Drinking Waters of the United States. *Sci. Total Environ.* **2017**, *581–582*, 909–922. https://doi.org/10.1016/j.scitotenv.2016.12.004.

(11) Steffen, W.; Richardson, K.; Rockström, J.; Cornell, S. E.; Fetzer, I.; Bennett, E. M.; Biggs, R.; Carpenter, S. R.; Vries, W. de; Wit, C. A. de; Folke, C.; Gerten, D.; Heinke, J.; Mace, G. M.; Persson, L. M.; Ramanathan, V.; Reyers, B.; Sörlin, S. Planetary Boundaries: Guiding Human Development on a Changing Planet. *Science (80-. ).* **2015**, *347*. https://doi.org/10.1126/science.1259855.

(12) Ahrens, L.; Bundschuh, M. Fate and Effects of Poly- and Perfluoroalkyl Substances in the Aquatic Environment: A Review. *Environ. Toxicol. Chem.* **2014**, *33* (9), 1921–1929. https://doi.org/10.1002/etc.2663.

(13) Bradley, P. M.; Kolpin, D. W.; Romanok, K. M.; Smalling, K. L.; Focazio, M. J.; Brown, J. B.; Cardon, M. C.; Carpenter, K. D.; Corsi, S. R.; Decicco, L. A.; et al. Reconnaissance of Mixed Organic and Inorganic Chemicals in Private and Public Supply Tapwaters at Selected Residential and Workplace Sites in the United States. *Environ. Sci. Technol.* **2018**, *52* (23), 13972–13985. https://doi.org/10.1021/acs.est.8b04622.

(14) Stamm, C.; Räsänen, K.; Burdon, F. J.; Altermatt, F.; Jokela, J.; Joss, A.; Ackermann, M.; Eggen, R. I. L. Unravelling the Impacts of Micropollutants in Aquatic Ecosystems: Interdisciplinary Studies at the Interface of Large-Scale Ecology. *Adv. Ecol. Res.* **2016**, *55*, 183–223. https://doi.org/10.1016/bs.aecr.2016.07.002.

(15) McCarroll, N. E.; Protzel, A.; Ioannou, Y.; Frank Stack, H.; Jackson, M. A.; Waters, M. D.; Dearfield, K. L. A Survey of EPA/OPP and Open Literature on Selected Pesticide Chemicals - III. Mutagenicity and Carcinogenicity of Benomyl and Carbendazim. *Mutat. Res. - Rev. Mutat. Res.* **2002**, *512* (1), 1–35. https://doi.org/10.1016/S1383-5742(02)00026-1.

(16) Altenburger, R.; Backhaus, T.; Boedeker, W.; Faust, M.; Scholze, M. Simplifying Complexity: Mixture Toxicity Assessment in the Last 20 Years. *Environ. Toxicol. Chem.* **2013**, *32* (8), 1685–1687. https://doi.org/10.1002/etc.2294.

(17) McKinlay, R.; Plant, J. A.; Bell, J. N. B.; Voulvoulis, N. Endocrine Disrupting Pesticides: Implications for Risk Assessment. *Environ. Int.* **2008**, *34* (2), 168–183. https://doi.org/10.1016/j.envint.2007.07.013.

(18) Colborn, T.; Vom Saal, F. S.; Soto, A. M. Developmental Effects of Endocrine-Disrupting Chemicals in Wildlife and Humans. *Environ. Health Perspect.* **1993**, *101* (5), 378–384. https://doi.org/10.1289/ehp.93101378.

(19) Hayes, T.; Haston, K.; Tsui, M.; Hoang, A.; Haeffele, C. Feminization of Male Frogs in the Wild Brief Communications. *Nature* **2002**, *419*, 895–896.

(20) Kidd, K. A.; Blanchfield, P. J.; Mills, K. H.; Palace, V. P.; Evans, R. E.; Lazorchak, J. M.; Flick, R. W. Collapse of a Fish Population after Exposure to a Synthetic Estrogen. *Proc. Natl. Acad. Sci.* **2007**, *104* (21), 8897–8901. https://doi.org/10.1073/pnas.0609568104.

(21) Peter, K. T.; Tian, Z.; Wu, C.; Lin, P.; White, S.; Du, B.; McIntyre, J. K.; Scholz, N. L.; Kolodziej, E. P. Using High-Resolution Mass Spectrometry to Identify Organic Contaminants Linked to Urban Stormwater Mortality Syndrome in Coho Salmon. *Environ. Sci. Technol.* **2018**, *52*, 10317–10327. https://doi.org/10.1021/acs.est.8b03287.

(22) Barry, V.; Winquist, A.; Steenland, K. Perfluorooctanoic Acid (PFOA) Exposures and Incident Cancers among Adults Living near a Chemical Plant. *Environ. Health Perspect.* **2013**, *121* (11–12), 1313–1318.

(23) U.S. EPA. Chemical Contaminants - CCL 4. *Environ. Proctection Agency* **2016**.

(24) U.S. EPA. Drinking Water Contaminants. *Environ. Proctection Agency* **2016**.

(25) U.S. EPA. Toxic Substances Control Act. *Environ. Proctection Agency* **2017**.

(26) U.S. EPA. Index of Chemical Names & Pesticide Chemical Codes Report. *Environ. Proctection Agency* **2016**.

(27) U.S. FDA. Resources for Information on Approved Drugs. *Fed. Drug Adm.* **2017**.

(28) Bletsou, A. A.; Jeon, J.; Hollender, J.; Archontaki, E.; Thomaidis, N. S. Targeted and Non-Targeted Liquid Chromatography-Mass Spectrometric Workflows for Identification of Transformation Products of Emerging Pollutants in the Aquatic Environment. *TrAC - Trends Anal. Chem.* **2015**, *66*, 32–44. https://doi.org/10.1016/j.trac.2014.11.009.

(29) Dolan, T.; Howsam, P.; Parsons, D. J.; Whelan, M. J. Is the EU Drinking Water Directive Standard for Pesticides in Drinking Water Consistent with the Precautionary Principle? *Environ. Sci. Technol.* **2013**, *47* (10), 4999–5006. https://doi.org/10.1021/es304955g.

(30) Kaboré, H. A.; Vo Duy, S.; Munoz, G.; Méité, L.; Desrosiers, M.; Liu, J.; Sory, T. K.; Sauvé, S. Worldwide Drinking Water Occurrence and Levels of Newly-Identified Perfluoroalkyl and Polyfluoroalkyl Substances. *Sci. Total Environ.* **2017**. https://doi.org/10.1016/j.scitotenv.2017.10.210.

(31) Olsen, G. W.; Lange, C. C.; Ellefson, M. E.; Mair, D. C.; Church, T. R.; Goldberg, C. L.; Herron, R. M.; Medhdizadehkashi, Z.; Nobiletti, J. B.; Rios, J. A.; Reagen, W. K.; Zobel, L. R. Temporal Trends of Perfluoroalkyl Concentrations in American Red Cross Adult Blood Donors, 2000-2010. *Environ. Sci. Technol.* **2012**, *46* (11), 6330–6338. https://doi.org/10.1021/es300604p.

(32) Bradley, P. M.; Journey, C. A.; Romanok, K. M.; Barber, L. B.; Buxton, H. T.; Foreman, W. T.; Furlong, E. T.; Glassmeyer, S. T.; Hladik, M. L.; Iwanowicz, L. R.; et al. Expanded Target-Chemical Analysis Reveals Extensive Mixed-Organic-Contaminant Exposure in U.S. Streams. *Environ. Sci. Technol.* **2017**, *51* (9), 4792–4802. https://doi.org/10.1021/acs.est.7b00012.

(33) Moschet, C.; Wittmer, I.; Simovic, J.; Junghans, M.; Piazzoli, A.; Singer, H.; Stamm, C.; Leu, C.; Hollender, J. How a Complete Pesticide Screening Changes the Assessment of Surface Water Quality. *Environ. Sci. Technol.* **2014**, *48* (10), 5423–5432. https://doi.org/10.1021/es500371t.

(34) Hollender, J.; Schymanski, E. L.; Singer, H.; Ferguson, P. L. Non-Target Screening with High Resolution Mass Spectrometry in the Environment: Ready to Go? *Environ. Sci. Technol.* **2017**, *51* (20), 11505–11512. https://doi.org/10.1021/acs.est.7b02184.

(35) Fairbairn, D. J.; Arnold, W. A.; Barber, B. L.; Kaufenberg, E. F.; Koskinen, W. C.; Novak, P. J.; Rice, P. J.; Swackhamer, D. L. Contaminants of Emerging Concern: Mass Balance and Comparison of Wastewater Effluent and Upstream Sources in a Mixed-Use Watershed. *Environ. Sci. Technol.* **2016**, *50* (1), 36–45. https://doi.org/10.1021/acs.est.5b03109.

(36) Zhang, X.; Lohmann, R.; Dassuncao, C.; Hu, X. C.; Weber, A. K.; Vecitis, C. D.; Sunderland, E. M. Source Attribution of Poly- and Perfluoroalkyl Substances (PFASs) in Surface Waters from Rhode Island and the New York Metropolitan Area. *Environ. Sci. Technol. Lett.* **2016**, *3* (9), 316–321. https://doi.org/10.1021/acs.estlett.6b00255.

(37) Ruff, M.; Mueller, M. S.; Loos, M.; Singer, H. P. 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 Res.* **2015**, *87*, 145–154. https://doi.org/10.1016/j.watres.2015.09.017.

(38) Ort, C.; Lawrence, M. G.; Rieckermann, J.; Joss, A. Sampling for Pharmaceuticals and Personal Care Products (PPCPs) and Illicit Drugs in Wastewater Systems: Are Your Conclusions Valid? A Critical Review. *Environ. Sci. Technol.* **2010**, *44* (16), 6024–6035. https://doi.org/10.1021/es100779n.

(39) Karpuzcu, M. E.; Fairbairn, D.; Arnold, W. A.; Barber, B. L.; Kaufenberg, E.; Koskinen, W. C.; Novak, P. J.; Rice, P. J.; Swackhamer, D. L. Identifying Sources of Emerging Organic Contaminants in a Mixed Use Watershed Using Principal Components Analysis. *Environ. Sci. Process. Impacts* **2014**, *16* (10), 2390–2399. https://doi.org/10.1039/c4em00324a.

(40) Munz, N. A.; Burdon, F. J.; de Zwart, D.; Junghans, M.; Melo, L.; Reyes, M.; Schonenberger, U.; Singer, H. P.; Spycher, B.; Hollender, J.; Stamm, C. Pesticides Drive Risk of Micropollutants in Wastewater-Impacted Streams during Low Flow Conditions. *Water Res.* **2017**, *110*, 366–377. https://doi.org/10.1016/j.watres.2016.11.001.

(41) Chiaia-Hernandez, A. C.; Günthardt, B. F.; Frey, M. P.; Hollender, J. Unravelling Contaminants in the Anthropocene Using Statistical Analysis of Liquid Chromatography−high-Resolution Mass Spectrometry Nontarget Screening Data Recorded in Lake Sediments. *Environ. Sci. Technol.* **2017**, *51* (21), 12547–12556. https://doi.org/10.1021/acs.est.7b03357.

(42) Wang, Y.; Lai, A.; Latino, D.; Fenner, K.; Helbling, D. E. Evaluating the Environmental Parameters That Determine Aerobic Biodegradation Half-Lives of Pesticides in Soil with a Multivariable Approach. *Chemosphere* **2018**, *209*, 430–438. https://doi.org/10.1016/j.chemosphere.2018.06.077.

(43) Ling, Y.; Klemes, M.; Steinschneider, S.; Dichtel, W. R.; Helbling, D. E. QSARs to Predict Adsorption Affinity of Organic Micropollutants for Activated Carbon and β-Cyclodextrin Polymer Adsorbents. *Submitt. to Water Res. Oct. 2018* **2019**, *154*, 217–226. https://doi.org/S0043135419301368.

(44) Veach, A. M.; Bernot, M. J. Temporal Variation of Pharmaceuticals in an Urban and Agriculturally Influenced Stream. *Sci. Total Environ.* **2011**, *409* (21), 4553–4563. https://doi.org/10.1016/j.scitotenv.2011.07.022.

(45) Pochodylo, A. L.; Helbling, D. E. Emerging Investigators Series: Prioritization of Suspect Hits in a Sensitive Suspect Screening Workflow for Comprehensive Micropollutant Characterization in Environmental Samples. *Environ. Sci. Water Res. Technol.* **2017**, *3* (1), 54–65. https://doi.org/10.1039/c6ew00248j.

(46) Focazio, M. J.; Kolpin, D. W.; Barnes, K. K.; Furlong, E. T.; Meyer, M. T.; Zaugg, S. D.; Barber, L. B.; Thurman, M. E. A National Reconnaissance for Pharmaceuticals and Other Organic Wastewater Contaminants in the United States - II) Untreated Drinking Water Sources. *Sci. Total Environ.* **2008**, *402* (2–3), 201–216. https://doi.org/10.1016/j.scitotenv.2008.02.021.

(47) Oulton, R. L.; Kohn, T.; Cwiertny, D. M. Pharmaceuticals and Personal Care Products in Effluent Matrices: A Survey of Transformation and Removal during Wastewater Treatment and Implications for Wastewater Management. *J. Environ. Monit.* **2010**, *12* (11), 1956–1978. https://doi.org/10.1039/c0em00068j.

(48) Nell, M.; Helbling, D. E. Exploring Matrix Effects and Quantifying Organic Additives in Hydraulic Fracturing Associated Fluids Using Liquid Chromatography Electrospray Ionization Mass Spectrometry. *Environ. Sci. Process. Impacts* **2018**, 195–205. https://doi.org/10.1039/c8em00135a.

(49) Kaufmann, A. The Current Role of High-Resolution Mass Spectrometry in Food Analysis. *Anal. Bioanal. Chem.* **2012**, *403* (5), 1233–1249. https://doi.org/10.1007/s00216-011-5629-4.

(50) Andra, S. S.; Austin, C.; Patel, D.; Dolios, G.; Awawda, M.; Arora, M. Trends in the Application of High-Resolution Mass Spectrometry for Human Biomonitoring: An Analytical Primer to Studying the Environmental Chemical Space of the Human Exposome. *Environ. Int.* **2017**, *100*, 32–61. https://doi.org/10.1016/j.envint.2016.11.026.

(51) Kasike, D.; Klinkmuller, K. D.; Sonneborn, M. Application of High-Performance Liquid Chromatography to Water Pollution Analysis. *J. Chromatogr. A* **1978**, *149*, 703–710. https://doi.org/10.1177/1077546310371491.

(52) Schymanski, E. L.; Singer, H. P.; Slobodnik, J.; Ipolyi, I. M.; Oswald, P.; Krauss, M.; Schulze, T.; Haglund, P.; Letzel, T.; Grosse, S.; et al. Non-Target Screening with High-Resolution Mass Spectrometry: Critical Review Using a Collaborative Trial on Water Analysis. *Anal. Bioanal. Chem.* **2015**, *407* (21), 6237–6255. https://doi.org/10.1007/s00216-015-8681-7.

(53) Loos, M. EnviMass Beta Version 3.4. *Zenodo*. 2016. https://doi.org/10.5281/zenodo.48164.

(54) Shaul, N. J.; Dodder, N. G.; Aluwihare, L. I.; Mackintosh, S. A.; Maruya, K. A.; Chivers, S. J.; Danil, K.; Weller, D. W.; Hoh, E. Nontargeted Biomonitoring of Halogenated Organic Compounds in Two Ecotypes of Bottlenose Dolphins (Tursiops Truncatus) from the Southern California Bight. *Environ. Sci. Technol.* **2015**, *49* (3), 1328–1338. https://doi.org/10.1021/es505156q.

(55) Moschet, C.; Anumol, T.; Lew, B. M.; Bennett, D. H.; Young, T. M. Household Dust as a Repository of Chemical Accumulation: New Insights from a Comprehensive High-Resolution Mass Spectrometric Study. *Environ. Sci. Technol.* **2018**, *52* (5), 2878–2887. https://doi.org/10.1021/acs.est.7b05767.

(56) Schymanski, E. L.; Singer, H. P.; Longrée, P.; Loos, M.; Ruff, M.; Stravs, M. A.; Ripollés Vidal, C.; Hollender, J. Strategies to Characterize Polar Organic Contamination in Wastewater: Exploring the Capability of High Resolution Mass Spectrometry. *Environ. Sci. Technol.* **2014**, *48* (3), 1811–1818. https://doi.org/10.1021/es4044374.

(57) Schlüsener, M. P.; Kunkel, U.; Ternes, T. A. Quaternary Triphenylphosphonium Compounds: A New Class of Environmental Pollutants. *Environ. Sci. Technol.* **2015**, *49* (24), 14282–14291. https://doi.org/10.1021/acs.est.5b03926.

(58) Sobus, J. R.; Wambaugh, J. F.; Isaacs, K. K.; Williams, A. J.; McEachran, A. D.; Richard, A. M.; Grulke, C. M.; Ulrich, E. M.; Rager, J. E.; Strynar, M. J.; Newton, S. R. Integrating Tools for Non-Targeted Analysis Research and Chemical Safety Evaluations at the US EPA. *J. Expo. Sci. Environ. Epidemiol.* **2017**, 1–16. https://doi.org/10.1038/s41370-017-0012-y.

(59) Zonja, B.; Delgado, A.; Pérez, S.; Barceló, D. LC-HRMS Suspect Screening for Detection-Based Prioritization of Iodinated Contrast Media Photodegradates in Surface Waters. *Environ. Sci. Technol.* **2015**, *49* (6), 3464–3472. https://doi.org/10.1021/es505250q.

(60) Thurman, E. M.; Ferrer, I.; Blotevogel, J.; Borch, T. Analysis of Hydraulic Fracturing Flowback and Produced Waters Using Accurate Mass: Identification of Ethoxylated Surfactants. *Anal. Chem.* **2014**, *86* (19), 9653–9661. https://doi.org/10.1021/ac502163k.

(61) Schymanski, E. L.; Jeon, J.; Gulde, R.; Fenner, K.; Ruff, M.; Singer, H. P.; Hollender, J. Identifying Small Molecules via High Resolution Mass Spectrometry: Communicating Confidence. *Environ. Sci. Technol.* **2014**, *48* (4), 2097–2098. https://doi.org/10.1021/es5002105 |.

(62) Royal Society of Chemistry. ChemSpider. 2018.

(63) Kim, S.; Thiessen, P. A.; Bolton, E. E.; Chen, J.; Fu, G.; Gindulyte, A.; Han, L.; He, J.; He, S.; Shoemaker, B. A.; Wang, J.; Yu, B.; Zhang, J.; Bryant, S. H. PubChem Substance and Compound Databases. *Nucleic Acids Res.* **2016**, *44* (D1), D1202–D1213. https://doi.org/10.1093/nar/gkv951.

(64) Ruttkies, C.; Schymanski, E. L.; Wolf, S.; Hollender, J.; Neumann, S. MetFrag Relaunched: Incorporating Strategies beyond in Silico Fragmentation. *J. Cheminform.* **2016**, *8* (1), 1–16. https://doi.org/10.1186/s13321-016-0115-9.

(65) Stravs, M. A.; Schymanski, E. L.; Singer, H. P.; Hollender, J. Automatic Recalibration and Processing of Tandem Mass Spectra Using Formula Annotation. *J. Mass Spectrom.* **2013**, *48* (1), 89–99. https://doi.org/10.1002/jms.3131.

(66) Daughton, C. G.; Ternes, T. A. Pharmaceuticals and Personal Care Products in the Environment: Agents of Subtle Change? *Environ. Health Perspect.* **1999**, *107* (Supplement 6), 907–938. https://doi.org/10.2307/3434573.

(67) Schmidt, T. C. Recent Trends in Water Analysis Triggering Future Monitoring of Organic Micropollutants. *Anal. Bioanal. Chem.* **2018**, 1–9. https://doi.org/10.1007/s00216-018-1015-9.

(68) Richardson, S. D.; Ternes, T. A. Water Analysis: Emerging Contaminants and Current Issues. *Anal. Chem.* **2014**, *86* (6), 2813–2848. https://doi.org/10.1021/ac500508t.

(69) Gómez, M. J.; Martínez Bueno, M. J.; Lacorte, S.; Fernández-Alba, A. R.; Agüera, A. Pilot Survey Monitoring Pharmaceuticals and Related Compounds in a Sewage Treatment Plant Located on the Mediterranean Coast. *Chemosphere* **2007**, *66* (6), 993–1002. https://doi.org/10.1016/j.chemosphere.2006.07.051.

(70) Chiaia, A. C.; Banta-Green, C.; Field, J. Eliminating Solid Phase Extraction with Large-Volume Injection LC / MS / MS : Analysis of Illicit and Legal Drugs and Human Urine Indicators in US Wastewaters Eliminating Solid Phase Extraction with Large-Volume Injection LC / MS / MS : Analysis of Illici. **2008**, *I* (23), 8841–8848. https://doi.org/10.1021/es802309v.

(71) Moschet, C.; Götz, C.; Longrée, P.; Hollender, J.; Singer, H. Multi-Level Approach for the Integrated Assessment of Polar Organic Micropollutants in an International Lake Catchment: The Example of Lake Constance. *Environ. Sci. Technol.* **2013**, *47* (13), 7028–7036. https://doi.org/10.1021/es304484w.

(72) Murray, K. E.; Thomas, S. M.; Bodour, A. A. Prioritizing Research for Trace Pollutants and Emerging Contaminants in the Freshwater Environment. *Environ. Pollut.* **2010**, *158* (12), 3462–3471. https://doi.org/10.1016/j.envpol.2010.08.009.

(73) Hajj-Mohamad, M.; Aboulfadl, K.; Darwano, H.; Madoux-Humery, A.-S.; Guérineau, H.; Sauvé, S.; Prévost, M.; Dorner, S. Wastewater Micropollutants as Tracers of Sewage Contamination: Analysis of Combined Sewer Overflow and Stream Sediments. *Environ. Sci. Process. Impacts* **2014**, *16* (10), 2442–2450. https://doi.org/10.1039/C4EM00314D.

(74) Brown, C. D.; van Beinum, W. Pesticide Transport via Sub-Surface Drains in Europe. *Environ. Pollut.* **2009**, *157* (12), 3314–3324. https://doi.org/10.1016/j.envpol.2009.06.029.

(75) Wittmer, I. K.; Bader, H. P.; Scheidegger, R.; Singer, H.; Lück, A.; Hanke, I.; Carlsson, C.; Stamm, C. Significance of Urban and Agricultural Land Use for Biocide and Pesticide Dynamics in Surface Waters. *Water Res.* **2010**, *44* (9), 2850–2862. https://doi.org/10.1016/j.watres.2010.01.030.

(76) Pal, A.; He, Y.; Jekel, M.; Reinhard, M.; Gin, K. Y. H. Emerging Contaminants of Public Health Significance as Water Quality Indicator Compounds in the Urban Water Cycle. *Environ. Int.* **2014**, *71*, 46–62. https://doi.org/10.1016/j.envint.2014.05.025.

(77) Cantwell, M. G.; Katz, D. R.; Sullivan, J. C.; Shapley, D.; Lipscomb, J.; Epstein, J.; Juhl, A. R.; Knudson, C.; O'Mullan, G. D. Spatial Patterns of Pharmaceuticals and Wastewater Tracers in the Hudson River Estuary. *Water Res.* **2018**, *137*, 335–343. https://doi.org/10.1016/j.watres.2017.12.044.

(78) Riverkeeper. Riverkeeper - NY's Clean Water Advocate.

(79) U.S. Geological Survey. National Water Information System data www.waterdata.usgs.gov/nwis.

(80) Geyer, W. R.; Chant, R. The Physical Oceanography Processes in the Hudson River Estuary. In *The Hudson River Estuary*; 2006; pp 24–38.

(81) Pochodylo, A. L.; Helbling, D. E. Emerging Investigators Series: Prioritization of Suspect Hits in a Sensitive Suspect Screening Workflow for Comprehensive Micropollutant Characterization in Environmental Samples. *Environ. Sci. Water Res. Technol.* **2017**, *3* (1), 54–65. https://doi.org/10.1039/c6ew00248j.

(82) Helbling, D. E.; Hollender, J.; Kohler, H. P. E.; Singer, H.; Fenner, K. High-Throughput Identification of Microbial Transformation Products of Organic Micropollutants. *Environ. Sci. Technol.* **2010**, *44* (17), 6621–6627. https://doi.org/10.1021/es100970m.

(83) Altenburger, R.; Ait-Aissa, S.; Antczak, P.; Backhaus, T.; Barceló, D.; Seiler, T. B.; Brion, F.; Busch, W.; Chipman, K.; de Alda, M. L.; et al. Future Water Quality Monitoring - Adapting Tools to Deal with Mixtures of Pollutants in Water Resource Management. *Sci. Total Environ.* **2015**, *512–513*, 540–551. https://doi.org/10.1016/j.scitotenv.2014.12.057.

(84) Yang, Y. Y.; Liu, W. R.; Liu, Y. S.; Zhao, J. L.; Zhang, Q. Q.; Zhang, M.; Zhang, J. N.; Jiang, Y. X.; Zhang, L. J.; Ying, G. G. Suitability of Pharmaceuticals and Personal Care Products (PPCPs) and Artificial Sweeteners (ASs) as Wastewater Indicators in the Pearl River Delta, South China. *Sci. Total Environ.* **2017**, *590–591*, 611–619. https://doi.org/10.1016/j.scitotenv.2017.03.001.

(85) U.S. EPA. Registration Eligibility Decision Thiophanate-Methyl. *U.S. Environ. Prot. Agency* **2005**.

(86) Chiaia-Hernandez, A. C.; Keller, A.; Wächter, D.; Steinlin, C.; Camenzuli, L.; Hollender, J.; Krauss, M. Long-Term Persistence of Pesticides and TPs in Archived Agricultural Soil Samples and Comparison with Pesticide Application. *Environ. Sci. Technol.* **2017**, *51* (18), 10642–10651. https://doi.org/10.1021/acs.est.7b02529.

(87) Nowell, L. H.; Moran, P. W.; Schmidt, T. S.; Norman, J. E.; Nakagaki, N.; Shoda, M. E.; Mahler, B. J.; Van Metre, P. C.; Stone, W. W.; Sandstrom, M. W.; Hladik, M. L. Complex Mixtures of Dissolved Pesticides Show Potential Aquatic Toxicity in a Synoptic Study of Midwestern U.S. Streams. *Sci. Total Environ.* **2018**, *613–614*, 1469–1488. https://doi.org/10.1016/j.scitotenv.2017.06.156.

(88) Castiglioni, S.; Davoli, E.; Riva, F.; Palmiotto, M.; Camporini, P.; Manenti, A.; Zuccato, E. Mass Balance of Emerging Contaminants in the Water Cycle of a Highly Urbanized and Industrialized Area of Italy. *Water Res.* **2018**, *131*, 287–298. https://doi.org/10.1016/j.watres.2017.12.047.

(89) Gilliom, R. J. Pesticides in U.S. Streams and Groundwater. *Environ. Sci. Technol.* **2007**, *41* (10), 3408–3414. https://doi.org/10.1021/es072531u.

(90) Wang, Z.; Cousins, I. T.; Scheringer, M.; Hungerbühler, K. Fluorinated Alternatives to Long-Chain Perfluoroalkyl Carboxylic Acids (PFCAs), Perfluoroalkane Sulfonic Acids (PFSAs) and Their Potential Precursors. *Environ. Int.* **2013**, *60* (2013), 242–248. https://doi.org/10.1016/j.envint.2013.08.021.

(91) Taniyasu, S.; Yamashita, N.; Yamazaki, E.; Petrick, G.; Kannan, K. The Environmental Photolysis of Perfluorooctanesulfonate, Perfluorooctanoate, and Related Fluorochemicals. *Chemosphere* **2013**, *90* (5), 1686–1692. https://doi.org/10.1016/j.chemosphere.2012.09.065.

(92) Lim, F. Y.; Ong, S. L.; Hu, J. Recent Advances in the Use of Chemical Markers for Tracing Wastewater Contamination in Aquatic Environment: A Review. *Water (Switzerland)* **2017**, *9* (2). https://doi.org/10.3390/w9020143.

(93) Graham, W. H.; Graham, D. W.; Denoyelles, F.; Smith, V. H.; Larive, C. K.; Thurman, E. M. Metolachlor and Alachlor Breakdown Product Formation Patterns in Aquatic Field Mesocosms. *Environ. Sci. Technol.* **1999**, *33* (24), 4471–4476. https://doi.org/10.1021/es990686z.

(94) Huntscha, S.; Singer, H.; Canonica, S.; Schwarzenbach, R. P.; Fenner, K. Input Dynamics and Fate in Surface Water of the Herbicide Metolachlor and of Its Highly Mobile Transformation Product Metolachlor ESA. *Environ. Sci. Technol.* **2008**, *42* (15), 5507–5513. https://doi.org/10.1021/es800395c.

(95) Loos, R.; Gawlik, B. M.; Locoro, G.; Rimaviciute, E.; Contini, S.; Bidoglio, G. EU-Wide Survey of Polar Organic Persistent Pollutants in European River Waters. *Environ. Pollut.* **2009**, *157* (2), 561–568. https://doi.org/10.1016/j.envpol.2008.09.020.

(96) Hanamoto, S.; Nakada, N.; Yamashita, N.; Hanamoto, S. Source Estimation of Pharmaceuticals Based on Catchment Population and In-Stream Attenuation in Yodo River Watershed, Japan. *Sci. Total Environ.* **2018**, *615*, 964–971. https://doi.org/10.1016/j.scitotenv.2017.10.013.

(97) McEachran, A. D.; Hedgespeth, M. L.; Newton, S. R.; McMahen, R.; Strynar, M.; Shea, D.; Nichols, E. G. Comparison of Emerging Contaminants in Receiving Waters Downstream of a Conventional Wastewater Treatment Plant and a Forest-Water Reuse System. *Environ. Sci. Pollut. Res.* **2018**, *25* (13), 12451–12463. https://doi.org/10.1007/s11356-018-1505-5.

(98) Carpenter, C.; Helbling, D. E. Widespread Micropollutant Monitoring in the Hudson River Estuary Reveals Spatiotemporal Micropollutant Clusters and Their Sources. *Environ. Sci. Technol.* **2018**, *52* (11), 6187–6196. https://doi.org/10.1021/acs.est.8b00945.

(99) Szöcs, E.; Brinke, M.; Karaoglan, B.; Schäfer, R. B. Large Scale Risks from Agricultural Pesticides in Small Streams. *Environ. Sci. Technol.* **2017**, *51* (13), 7378–7385. https://doi.org/10.1021/acs.est.7b00933.

(100) Spycher, S.; Mangold, S.; Doppler, T.; Junghans, M.; Wittmer, I.; Stamm, C.; Singer, H. Pesticide Risks in Small Streams - How to Get as Close as Possible to the Stress Imposed on Aquatic Organisms. *Environ. Sci. Technol.* **2018**, *52* (8), 4526–4535. https://doi.org/10.1021/acs.est.8b00077.

(101) Lorenz, S.; Rasmussen, J. J.; Süß, A.; Kalettka, T.; Golla, B.; Horney, P.; Stähler, M.; Hommel, B.; Schäfer, R. B. Specifics and Challenges of Assessing Exposure and Effects of Pesticides in Small Water Bodies. *Hydrobiologia* **2017**, *793* (1), 213–224. https://doi.org/10.1007/s10750-016-2973-6.

(102) U.S. EPA. *Geographic Information Systems Analysis of the Surface Drinking Water Provided by Intermittent, Ephemeral, and Headwater Streams in the U.S.*; 2009.

(103) Neale, P. A.; Munz, N. A.; Aït-Aïssa, S.; Altenburger, R.; Brion, F.; Busch, W.; Escher, B. I.; Hilscherovïcá, K.; Kienle, C.; Novák, J.; Seiler, T. B.; Shao, Y.; Stamm, C.; Hollender, J. Integrating Chemical Analysis and Bioanalysis to Evaluate the Contribution of Wastewater Effluent on the Micropollutant Burden in Small Streams. *Sci. Total Environ.* **2017**, *576*, 785–795. https://doi.org/10.1016/j.scitotenv.2016.10.141.

(104) Schollée, J. E.; Schymanski, E. L.; Avak, S. E.; Loos, M.; Hollender, J. Prioritizing Unknown Transformation Products from Biologically-Treated Wastewater Using High-Resolution Mass Spectrometry, Multivariate Statistics, and Metabolic Logic. *Anal. Chem.* **2015**, *87* (24), 12121–12129. https://doi.org/10.1021/acs.analchem.5b02905.

(105) Chiaia-Hernandez, A. C.; Schymanski, E. L.; Kumar, P.; Singer, H. P.; Hollender, J. Suspect and Nontarget Screening Approaches to Identify Organic Contaminant Records in Lake Sediments. *Anal. Bioanal. Chem.* **2014**, *406* (28), 7323–7335. https://doi.org/10.1007/s00216-014-8166-0.

(106) Smith, C. A.; Want, E. J.; O'Maille, G.; Abagyan, R.; Siuzdak, G. XCMS: Processing Mass Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment, Matching, and Identification. *Anal. Chem.* **2006**, *78* (3), 779–787. https://doi.org/10.1021/ac051437y.

(107) Allen, F.; Pon, A.; Wilson, M.; Greiner, R.; Wishart, D. CFM-ID: A Web Server for Annotation, Spectrum Prediction and Metabolite Identification from Tandem Mass Spectra. *Nucleic Acids Res.* **2014**, *42* (W1), 94–99. https://doi.org/10.1093/nar/gku436.

(108) Pluskal, T.; Castillo, S.; Villar-Briones, A.; Orešič, M. MZmine 2: Modular Framework for Processing, Visualizing, and Analyzing Mass Spectrometry-Based Molecular Profile Data. *BMC Bioinformatics* **2010**, *11* (395), 1–11. https://doi.org/10.1186/1471-2105-11-395.

(109) HighChem. MzCloud. **2013**.

(110) Stein, S. E. Chemical Substructure Identification by Mass Spectral Library Searching. *J. Am. Soc. Mass Spectrom.* **1995**, *6* (8), 644–655. https://doi.org/10.1016/1044-0305(95)00291-K.

(111) Horai, H.; Arita, M.; Kanaya, S.; Nihei, Y.; Ikeda, T.; Suwa, K.; Ojima, Y.; Tanaka, K.; Tanaka, S.; Aoshima, K.; et al. MassBank: A Public Repository for Sharing Mass Spectral Data for Life Sciences. *J. Mass Spectrom.* **2010**, *45* (7), 703–714. https://doi.org/10.1002/jms.1777.

(112) Smith, C. A.; O'Maille, G.; Want, E. J.; Qin, C.; Trauger, S. A.; Brandon, T. R.; Custodio, D. E.; Abagyan, R.; Siuzdak, G. Metlin: A Metabolite Mass Spectral Database. In *Proceedings of the 9th International Congress of Therapeutic Drug Monitoring & Clinical Toxicology*; 2005; Vol. 27, pp 747–751.

(113) Ryberg, K. R.; Vecchia, A. V.; Gilliom, R. J.; Martin, J. D. *Pesticide Trends in Major Rivers of the United States, 1992-2010*; Scientific Investigations Report 2014-5135; U.S. Geological Survey, 2014. https://doi.org/10.3133/sir20145135.

(114) Richards, B. K.; Pacenka, S.; Meyer, M. T.; Dietze, J. E.; Schatz, A. L.; Teuffer, K.; Aristilde, L.; Steenhuis, T. S. Antecedent and Post-Application Rain Events Trigger Glyphosate Transport from Runoff-Prone Soils. *Environ. Sci. Technol. Lett.* **2018**, *5* (5), 249–254. https://doi.org/10.1021/acs.estlett.8b00085.

(115) Homer, C. G.; Dewitz, J. A.; Yang, L.; Jin, S.; Danielson, P.; Xian, G.; Coulston, J.; Herold, N. D.; Wickham, J. D.; Megown, K. Completion of the 2011 National Land Cover Database for the Conterminous United States-Representing a Decade of Land Cover Change Information. *Photogramm. Eng. Remote Sensing* **2015**, *81* (5), 345–354. https://doi.org/10.14358/PERS.81.5.345.

(116) U.S. EPA. EPA Facility Registry Service (FRS): Wastewater Treatment Plants www.catalog.data.gov/dataset/epa-facility-registry-service-frs-wastewater-treatment-plants.

(117) Hollingshead, N. *Residental Wastewater Systems; Cayuga Lake Watershed Network*; 2007.

(118) Carpenter, C. M. G.; Helbling, D. E. Removal of Micropollutants in Biofilters: Hydrodynamic Effects on Biofilm Assembly and Functioning. *Water Res.* **2017**, *120*, 211–221. https://doi.org/10.1016/j.watres.2017.04.071.

(119) Kessner, D.; Chambers, M.; Burke, R.; Agus, D.; Mallick, P. ProteoWizard: Open Source Software for Rapid Proteomics Tools Development. *Bioinformatics* **2008**, *24* (21), 2534–2536. https://doi.org/10.1093/bioinformatics/btn323.

(120) R Development Core Team. R: A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing*. Vienna, Austria 2008.

(121) Meringer, M.; Reinker, S.; Zhang, J.; Muller, A. MS/MS Data Improves Automated Determination of Molecular Formulas by Mass Spectrometry. *MATCH Commun Math Comput Chem* **2011**, *65* (2), 259–290.

(122) Gerlich, M.; Neumann, S. MetFusion: Integration of Compound Identification Strategies. *J. Mass Spectrom.* **2013**, *48* (3), 291–298. https://doi.org/10.1002/jms.3123.

(123) Norman network. Norman Suspect List Exchange http://www.norman-network.com/?q=node/236.

(124) Campanha, M. B.; Awan, A. T.; de Sousa, D. N. R.; Grosseli, G. M.; Mozeto, A. A.; Fadini, P. S. A 3-Year Study on Occurrence of Emerging Contaminants in an Urban Stream of São Paulo State of Southeast Brazil. *Environ. Sci. Pollut. Res.* **2015**, *22* (10), 7936–7947. https://doi.org/10.1007/s11356-014-3929-x.

(125) Osorio, V.; Proia, L.; Ricart, M.; Pérez, S.; Ginebreda, A.; Cortina, J. L.; Sabater, S.; Barceló, D. Hydrological Variation Modulates Pharmaceutical Levels and Biofilm Responses in a Mediterranean River. *Sci. Total Environ.* **2014**, *472* (2014), 1052–1061. https://doi.org/10.1016/j.scitotenv.2013.11.069.

(126) Kolpin, D. W.; Skopec, M.; Meyer, M. T.; Furlong, E. T.; Zaugg, S. D. Urban Contribution of Pharmaceuticals and Other Organic Wastewater Contaminants to Streams during Differing Flow Conditions. *Sci. Total Environ.* **2004**, *328* (1–3), 119–130. https://doi.org/10.1016/j.scitotenv.2004.01.015.

(127) Doppler, T.; Camenzuli, L.; Hirzel, G.; Krauss, M.; Lück, A.; Stamm, C. Spatial Variability of Herbicide Mobilisation and Transport at Catchment Scale: Insights from a Field Experiment. *Hydrol. Earth Syst. Sci.* **2012**, *16* (7), 1947–1967. https://doi.org/10.5194/hess-16-1947-2012.

(128) McEachran, A. D.; Sobus, J. R.; Williams, A. J. Identifying Known Unknowns Using the US EPA's CompTox Chemistry Dashboard. *Anal. Bioanal. Chem.* **2017**, *409* (7), 1729–1735. https://doi.org/10.1007/s00216-016-0139-z.

(129) Little, J. L.; Cleven, C.; Brown, S. Identification of "Known Unknowns" Utilizing Accurate Mass Data and Chemical Abstracts Service Databases. *J. Am. Soc. Mass Spectrom.* **2011**, *22* (1), 348–359. https://doi.org/10.1007/s13361-010-0034-3.

(130) Little, J. L.; Williams, A. J.; Pshenichnov, A.; Tkachenko, V. Identification of "Known Unknowns" Utilizing Accurate Mass Data and Chemspider. *J. Am. Soc. Mass Spectrom.* **2012**, *23* (1), 179–185. https://doi.org/10.1007/s13361-011-0265-y.

(131) Hu, M.; Müller, E.; Schymanski, E. L.; Ruttkies, C.; Schulze, T.; Brack, W.; Krauss, M. Performance of Combined Fragmentation and Retention Prediction for the Identification of Organic Micropollutants by LC-HRMS. *Anal. Bioanal. Chem.* **2018**, *410* (7), 1931–1941. https://doi.org/10.1007/s00216-018-0857-5.

(132) De Wever, H.; Besse, P.; Verachtert, H. Microbial Transformations of 2-Substituted Benzothiazoles. *Appl. Microbiol. Biotechnol.* **2001**, *57* (5–6), 620–625. https://doi.org/10.1007/s00253-001-0842-2.

(133) Liao, C.; Kim, U. J.; Kannan, K. A Review of Environmental Occurrence, Fate, Exposure, and Toxicity of Benzothiazoles. *Environ. Sci. Technol.* **2018**, *52* (9), 5007–5026. https://doi.org/10.1021/acs.est.7b05493.

(134) Grebel, J. E.; Mohanty, S. K.; Torkelson, A. A.; Boehm, A. B.; Higgins, C. P.; Maxwell, R. M.; Nelson, K. L.; Sedlak, D. L. Engineered Infiltration Systems for Urban Stormwater Reclamation. *Environ. Eng. Sci.* **2013**, *30* (8), 437–454. https://doi.org/10.1089/ees.2012.0312.

(135) Hug, C.; Ulrich, N.; Schulze, T.; Brack, W.; Krauss, M. Identification of Novel Micropollutants in Wastewater by a Combination of Suspect and Nontarget Screening. *Environ. Pollut.* **2014**, *184*, 25–32. https://doi.org/10.1016/j.envpol.2013.07.048.

(136) Blum, K. M.; Andersson, P. L.; Renman, G.; Ahrens, L.; Gros, M.; Wiberg, K.; Haglund, P. Non-Target Screening and Prioritization of Potentially Persistent, Bioaccumulating and Toxic Domestic Wastewater Contaminants and Their Removal in on-Site and Large-Scale Sewage Treatment Plants. *Sci. Total Environ.* **2017**, *575*, 265–275. https://doi.org/10.1016/j.scitotenv.2016.09.135.

(137) Baken, K. A.; Sjerps, R. M. A.; Schriks, M.; van Wezel, A. P. Toxicological Risk Assessment and Prioritization of Drinking Water Relevant Contaminants of Emerging Concern. *Environ. Int.* **2018**, *118* (May), 293–303. https://doi.org/10.1016/j.envint.2018.05.006.

(138) Unice, K. M.; Bare, J. L.; Kreider, M. L.; Panko, J. M. Experimental Methodology for Assessing the Environmental Fate of Organic Chemicals in Polymer Matrices Using Column Leaching Studies and OECD 308 Water/Sediment Systems: Application to Tire and Road Wear Particles. *Sci. Total Environ.* **2015**, *533*, 476–487. https://doi.org/10.1016/j.scitotenv.2015.06.053.

(139) Zeng, F.; Wen, J.; Cui, K.; Wu, L.; Liu, M.; Li, Y.; Lin, Y.; Zhu, F.; Ma, Z.; Zeng, Z. Seasonal Distribution of Phthalate Esters in Surface Water of the Urban Lakes in the Subtropical City, Guangzhou, China. *J. Hazard. Mater.* **2009**, *169* (1–3), 719–725. https://doi.org/10.1016/j.jhazmat.2009.04.006.

(140) Kim, U. J.; Kannan, K. Occurrence and Distribution of Organophosphate Flame Retardants/Plasticizers in Surface Waters, Tap Water, and Rainwater: Implications for Human Exposure. *Environ. Sci. Technol.* **2018**, *52* (10), 5625–5633. https://doi.org/10.1021/acs.est.8b00727.

(141) Wei, G. L.; Li, D. Q.; Zhuo, M. N.; Liao, Y. S.; Xie, Z. Y.; Guo, T. L.; Li, J. J.; Zhang, S. Y.; Liang, Z. Q. Organophosphorus Flame Retardants and Plasticizers: Sources, Occurrence, Toxicity and Human Exposure. *Environ. Pollut.* **2015**, *196*, 29–46. https://doi.org/10.1016/j.envpol.2014.09.012.

(142) Huang, F.; Zou, S.; Deng, D.; Lang, H.; Liu, F. Antibiotics in a Typical Karst River System in China: Spatiotemporal Variation and Environmental Risks. *Sci. Total Environ.* **2019**, *650*, 1348–1355. https://doi.org/10.1016/j.scitotenv.2018.09.131.

(143) Schwarzenbach, R. P.; Gschwend, P. M.; Imboden, D. M. *Environmental Organic Chemistry*, Third.; Wiley, 2002. https://doi.org/10.1002/0471649643.

(144) Mahler, B. J.; Van Metre, P. C.; Burley, T. E.; Loftin, K. A.; Meyer, M. T.; Nowell, L. H. Similarities and Differences in Occurrence and Temporal Fluctuations in Glyphosate and Atrazine in Small Midwestern Streams (USA) during the 2013 Growing Season. *Sci. Total Environ.* **2017**, *579*, 149–158. https://doi.org/10.1016/j.scitotenv.2016.10.236.

(145) Parajulee, A.; Lei, Y. D.; De Silva, A. O.; Cao, X.; Mitchell, C. P. J.; Wania, F. Assessing the Source-to-Stream Transport of Benzotriazoles during Rainfall and Snowmelt in Urban and Agricultural Watersheds. *Environ. Sci. Technol.* **2017**, *51* (8), 4191–4198. https://doi.org/10.1021/acs.est.6b05638.

(146) Carpenter, C.; Wong, L. Y. J.; Johnson, C. A.; Helbling, D. E. Fall Creek Monitoring Station: Highly Resolved Temporal Sampling to Prioritize the Identification of Nontarget Micropollutants in a Small Stream. *Environ. Sci. Technol.* **2019**, *53* (1), 77–87. https://doi.org/10.1021/acs.est.8b05320.

(147) Vogel, R. M.; Fennessey, N. M. Flow Duration Curves II: A Review of Applications in Water Resources Planning. *J. Am. Water Resour. Assoc.* **1995**, *31* (6), 1029–1039.

(148) U.S. EPA. iCSS ToxCast Dashboard https://comptox.epa.gov/dashboard/.

(149) De Cicco, L. A.; Corsi, S. R.; Villeneuve, D. L.; Blackwell, B. .; Ankley, G. T. ToxEval: Evaluation of Measured Concentration Data Using the ToxCast High-Throughput Screening Database or a User-Defined Set of Concentration Benchmarks. **2018**. https://doi.org/doi:10.5066/P906UQ5I.

(150) Becker, R. A.; Friedman, K. P.; Simon, T. W.; Marty, M. S.; Patlewicz, G.; Rowlands, J. C. An Exposure: Activity Profiling Method for Interpreting High-Throughput Screening Data for Estrogenic Activity-Proof of Concept. *Regul. Toxicol. Pharmacol.* **2015**, *71* (3), 398–408. https://doi.org/10.1016/j.yrtph.2015.01.008.

(151) Li, S.; Villeneuve, D. L.; Berninger, J. P.; Blackwell, B. R.; Cavallin, J. E.; Hughes, M. N.; Jensen, K. M.; Kahl, M. D.; Stevens, K. E.; Thomas, L. M.; Weberg, M. A.; Ankley, G. T.; Li, S.; Berninger, J. P.; Jorgenson, Z.; Schroeder, A. L. An Integrated Approach for Identifying Priority Contaminant in the Great Lakes Basin – Investigations in the Lower Green Bay/Fox River and Milwaukee Estuary Areas of Concern. *Sci. Total Environ.* **2017**, *579*, 825–837. https://doi.org/10.1016/j.scitotenv.2016.11.021.

(152) Blackwell, B. R.; Ankley, G. T.; Corsi, S. R.; Decicco, L. A.; Houck, K. A.; Judson, R. S.; Li, S.; Martin, M. T.; Murphy, E.; Schroeder, A. L.; Smith, E. R.; Swintek, J.; Villeneuve, D. L. An "EAR" on Environmental Surveillance and Monitoring: A Case Study on the Use of Exposure-Activity Ratios (EARs) to Prioritize Sites, Chemicals, and Bioactivities of Concern in Great Lakes Waters. *Environ. Sci. Technol.* **2017**, *51* (15), 8713–8724. https://doi.org/10.1021/acs.est.7b01613.

(153) Menne, M. J.; Durre, I.; Korzeniewski, B.; McNeal, S.; Thomas, K.; Yin, X.; Anthony, S.; Ray, R.; Vose, R. S.; E.Gleason, B.; Houston, T. G. Global Historical Climatology Network - Daily (GHCN-Daily), Version 3. *NOAA Natl. Clim. Data Cent.* **2012**. https://doi.org/10.7289/V5D21VHZ.

(154) Chamberlain, S. rnoaa: "NOAA" Weather Data from R https://cran.r-project.org/package=rnoaa.

(155) NOAA - Northeast Regional Climate Center at Cornell University. Climate Data. 2018.

(156) Buuren, S. van; Groothuis-Oudshoorn, K. Mice: Multivariate Imputation by Chained Equations in R. *J. Stat. Softw.* **2011**, *45* (3). https://doi.org/10.18637/jss.v045.i03.

(157) Esbensen, K. H.; Guyot, D.; Westad, F.; Houmoller, L. P. *Multivariate Data Analysis-In Practice: An Introduction to Multivariate Data Analysis and Experimental Design*, 5th ed.; CAMO Prcoess AS: Olso, Norway, 2002.

(158) Lee, K. J.; Carlin, J. B. Multiple Imputation in the Presence of Non-Normal Data. *Stat. Med.* **2017**, *36* (4), 606–617. https://doi.org/10.1002/sim.7173.

(159) Sterne, J. A. C.; White, I. R.; Carlin, J. B.; Spratt, M.; Royston, P.; Kenward, M. G.; Wood, A. M.; Carpenter, J. R. Multiple Imputation for Missing Data in Epidemiological and Clinical Research: Potential and Pitfalls. *Br. Med. J.* **2009**, *338* (b2393). https://doi.org/10.1136/bmj.b2393.

(160) Long, J. A. Jtools: Analysis and Presentation of Social ScientificData. *R Packag.* **2018**.

(161) Hyndman, R. J.; Khandakar, Y. Automatic Time Series Forecasting: The Forecast Package for R. *J. Stat. Softw.* **2008**, *27* (3), 1–22.

(162) Pinheiro, J.; Bates, D.; DebRoy, S.; Sarkar, D.; R Core Team. Nlme: Linear and Nonlinear Mixed Effects Models. *R Packag.* **2018**.

(163) Pfaff, B. *Analysis of Integrated and Cointegrated Time Series with R*, 2nd ed.; Springer: New York, 2008.

(164) Calcagno, V.; Mazancourt, C. de. Glmulti: An R Package for Easy Automated Model Selection with (Generalized) Linear Models. *J. Stat. Softw.* **2010**, *34* (12). https://doi.org/10.18637/jss.v034.i12.

(165) Zuur, A. F.; Ieno, E. N.; Elphick, C. S. A Protocol for Data Exploration to Avoid Common Statistical Problems. *Methods Ecol. Evol.* **2010**, *1* (1), 3–14. https://doi.org/10.1111/j.2041-210X.2009.00001.x.

(166) Burnham, K. P.; Anderson, D. R. *Model Selection and Multimodel Inference A Practical Information-Theoretic Approach*, 2nd ed.; Springer, 2002. https://doi.org/10.1360/zd-2013-43-6-1064.

(167) Barton, K. Multi-Model Inference (MuMIn). *R Packag.* **2018**. https://doi.org/Available at www.cran.r-project.org/web/packages.

(168) Schomaker, M.; Heumann, C. Model Selection and Model Averaging after Multiple Imputation. *Comput. Stat. Data Anal.* **2014**, *71*, 758–770. https://doi.org/10.1016/j.csda.2013.02.017.

(169) Racine, J. Consistent Cross-Validatory Model-Selection for Dependent Data: Hv-Block Cross-Validation. *J. Econom.* **2000**, *99* (1), 39–61. https://doi.org/10.1016/S0304-4076(00)00030-0.

(170) Pampaka, M.; Hutcheson, G.; Williams, J. Handling Missing Data: Analysis of a Challenging Data Set Using Multiple Imputation. *Int. J. Res. Method Educ.* **2016**, *39* (1), 19–37. https://doi.org/10.1080/1743727X.2014.979146.

(171) Altmann, J.; Massa, L.; Sperlich, A.; Gnirss, R.; Jekel, M. UV254 Absorbance as Real-Time Monitoring and Control Parameter for Micropollutant Removal in Advanced Wastewater Treatment with Powdered Activated Carbon. *Water Res.* **2016**, *94*, 240–245. https://doi.org/10.1016/j.watres.2016.03.001.

(172) Baker, A. Fluorescence Excitation - Emission Matrix Characterization of Some Sewage-Impacted Rivers. *Environ. Sci. Technol.* **2001**, *35* (5), 948–953. https://doi.org/10.1021/es000177t.

(173) Lynch, J. A.; Hanna, C. M.; Corbett, E. S. Predicting PH, Alkalinity, and Total Acidity in Stream Water During Episodic Events. *Water Resour. Res.* **1986**, *22* (6), 905–912. https://doi.org/10.1029/WR022i006p00905.

(174) Caissie, D.; Pollock, T. L.; Cunjak, R. A. Variation in Stream Water Chemistry and Hydrograph Separation in a Small Drainage Basin. *J. Hydrol.* **1996**, *178* (1–4), 137–157. https://doi.org/10.1016/0022-1694(95)02806-4.

(175) Laird, D. A.; Yen, P. Y.; Koskinen, W. C.; Steinhelmer, T. R.; Dowdy, R. H. Sorption of Atrazine on Soil Clay Components. *Environ. Sci. Technol.* **1994**, *28* (6), 1054–1061. https://doi.org/10.1021/es00055a014.

(176) Naja, G. M.; Volesky, B. Adsorption and Desorption of Atrazine, Deethylatrazine, Deisopropylatrazine, Hydroxyatrazine, and Metolachlor in Two Soils from Virginia. *J. Environ. Qual.* **1996**, *25*, 1179–1185. https://doi.org/10.1128/AEM.70.5.2989.

(177) Barceló, D.; Cortina, J. L.; Marcé, R.; Ginebreda, A.; Osorio, V.; Pérez, S. Occurrence and Modeling of Pharmaceuticals on a Sewage-Impacted Mediterranean River and Their Dynamics under Different Hydrological Conditions. *Sci. Total Environ.* **2012**, *440*, 3–13. https://doi.org/10.1016/j.scitotenv.2012.08.040.

(178) Li, Z.; Sobek, A.; Radke, M. Fate of Pharmaceuticals and Their Transformation Products in Four Small European Rivers Receiving Treated Wastewater. *Environ. Sci. Technol.* **2016**, *50* (11), 5614–5621. https://doi.org/10.1021/acs.est.5b06327.

(179) Singer, H.; Radny, D.; Borer, P.; Rothardt, J.; Huggenberger, P.; Epting, J.; Hollender, J.; Loos, M. Comprehensive Micropollutant Screening Using LC-HRMS/MS at Three Riverbank Filtration Sites to Assess Natural Attenuation and Potential Implications for Human Health. *Water Res. X* **2018**, *1*, 100007. https://doi.org/10.1016/j.wroa.2018.100007.

(180) Kaserzon, S. L.; Heffernan, A. L.; Thompson, K.; Mueller, J. F.; Gomez Ramos, M. J. Rapid Screening and Identification of Chemical Hazards in Surface and Drinking Water Using High Resolution Mass Spectrometry and a Case-Control Filter. *Chemosphere* **2017**, *182*, 656–664. https://doi.org/10.1016/j.chemosphere.2017.05.071.

(181) Qi, W.; Müller, B.; Pernet-Coudrier, B.; Singer, H.; Liu, H.; Qu, J.; Berg, M. Organic Micropollutants in the Yangtze River: Seasonal Occurrence and Annual Loads. *Sci. Total Environ.* **2014**, *472*, 789–799. https://doi.org/10.1016/j.scitotenv.2013.11.019.

(182) Brack, W.; Ait-Aissa, S.; Burgess, R. M.; Busch, W.; Creusot, N.; Di Paolo, C.; Escher, B. I.; Mark Hewitt, L.; Hilscherova, K.; Hollender, J.; et al. Effect-Directed Analysis Supporting Monitoring of Aquatic Environments - An in-Depth Overview. *Sci. Total Environ.* **2016**, *544*, 1073–1118. https://doi.org/10.1016/j.scitotenv.2015.11.102.

(183) Huntscha, S.; Singer, H. P.; McArdell, C. S.; Frank, C. E.; Hollender, J. Multiresidue Analysis of 88 Polar Organic Micropollutants in Ground, Surface and Wastewater Using Online Mixed-Bed Multilayer Solid-Phase Extraction Coupled to High Performance Liquid Chromatography-Tandem Mass Spectrometry. *J. Chromatogr. A* **2012**, *1268*, 74–83. https://doi.org/10.1016/j.chroma.2012.10.032.

(184) Gros, M.; Rodríguez-Mozaz, S.; Barceló, D. Fast and Comprehensive Multi-Residue Analysis of a Broad Range of Human and Veterinary Pharmaceuticals and Some of Their Metabolites in Surface and Treated Waters by Ultra-High-Performance Liquid Chromatography Coupled to Quadrupole-Linear Ion Trap Tandem. *J. Chromatogr. A* **2012**, *1248*, 104–121. https://doi.org/10.1016/j.chroma.2012.05.084.

(185) Singer, H. P.; Wössner, A. E.; McArdell, C. S.; Fenner, K. Rapid Screening for Exposure to "Non-Target" Pharmaceuticals from Wastewater Effluents by Combining HRMS-Based Suspect Screening and Exposure Modeling. *Environ. Sci. Technol.* **2016**, *50* (13), 6698–6707. https://doi.org/10.1021/acs.est.5b03332.

(186) Zhou, J.; Broodbank, N. Sediment-Water Interactions of Pharmaceutical Residues in the River Environment. *Water Res.* **2014**, *48* (1), 61–70. https://doi.org/10.1016/j.watres.2013.09.026.

(187) Sosnin, S.; Karlov, D.; Tetko, I. V.; Fedorov, M. V. Comparative Study of Multitask Toxicity Modeling on a Broad Chemical Space. *J. Chem. Inf. Model.* **2018**, *59*, 1062–1072. https://doi.org/10.1021/acs.jcim.8b00685.

(188) Li, Y.; Zhu, G.; Ng, W. J.; Tan, S. K. A Review on Removing Pharmaceutical Contaminants from Wastewater by Constructed Wetlands: Design, Performance and Mechanism. *Sci. Total Environ.* **2014**, *468–469*, 908–932. https://doi.org/10.1016/j.scitotenv.2013.09.018.

(189) Hoppe-Jones, C.; Dickenson, E. R. V.; Drewes, J. E. The Role of Microbial Adaptation and Biodegradable Dissolved Organic Carbon on the Attenuation of Trace Organic Chemicals during Groundwater Recharge. *Sci. Total Environ.* **2012**, *437*, 137–144. https://doi.org/10.1016/j.scitotenv.2012.08.009.

(190) Peter, K. T.; Herzog, S.; Tian, Z.; Wu, C.; McCray, J. E.; Lynch, K.; Kolodziej, E. P. Evaluating Emerging Organic Contaminant Removal in an Engineered Hyporheic Zone Using High Resolution Mass Spectrometry. *Water Res.* **2019**, *150*, 140–152. https://doi.org/10.1016/j.watres.2018.11.050.

(191) Alsbaiee, A.; Smith, B. J.; Xiao, L.; Ling, Y.; Helbling, D. E.; Dichtel, W. R. Rapid Removal of Organic Micropollutants from Water by a Porous β-Cyclodextrin Polymer. *Nature* **2016**, *529* (7585), 190–194. https://doi.org/10.1038/nature16185.

(192) Margot, J.; Kienle, C.; Magnet, A.; Weil, M.; Rossi, L.; de Alencastro, L. F.; Abegglen, C.; Thonney, D.; Chèvre, N.; Schärer, M.; Barry, D. A. Treatment of Micropollutants in Municipal Wastewater: Ozone or Powdered Activated Carbon? *Sci. Total Environ.* **2013**, *461–462*, 480–498. https://doi.org/10.1016/j.scitotenv.2013.05.034.

(193) Lee, C. O.; Howe, K. J.; Thomson, B. M. Ozone and Biofiltration as an Alternative to Reverse Osmosis for Removing PPCPs and Micropollutants from Treated Wastewater. *Water Res.* **2012**, *46* (4), 1005–1014. https://doi.org/10.1016/j.watres.2011.11.069.

# APPENDICES

# APPENDICES LIST OF FIGURES

**Appendix A**



**Appendix B**

**Appendix C**

# APPENDICES LIST OF TABLES

**Appendix A**



**Appendix B**



**Appendix C**

# APPENDIX A – Widespread Micropollutant Monitoring in the Hudson River Estuary Reveals Spatiotemporal Micropollutant Clusters and Their Sources

## A.1 Study area

**Table A1:** Sample site IDs, names, and sampling locations along the Hudson River Estuary.

| Site ID | Site Name | River distance (km) | Type of Site | Sample site ID[a] |
|---|---|---|---|---|
| 1 | Orangetown STP Outfall | 41.8 | STP Outfall | O_STP |
| 2 | TZ Bridge | 44.3 | Mid-channel | LHR |
| 3 | Pocantico River | 45.1 | Tributary (mouth) | PR_M |
| 4 | Cedar Pond Brook | 64.4 | Tributary (mouth) | CB_M |
| 5 | Furnace Brook | 64.5 | Tributary (mouth) | FB_M |
| 6 | Annesville Creek | 70.8 | Tributary (mouth) | AC_M |
| 7 | West Point STP Outfall | 84.5 | STP Outfall | W_STP |
| 8 | Rondout Creek–Kingston STP Outfall | 148.2 | STP Outfall | R_STP |
| 9 | Rondout Creek–Kingston Public Dock | 148.4 | Tributary (inside) | RC_D |
| 10 | Rondout Creek–Eddyville Anchorage | 148.5 | Tributary (inside) | RC_U |
| 11 | Esopus Creek–Entrance | 164.2 | Tributary (inside) | EC_D |
| 12 | Esopus Creek–West | 164.3 | Tributary (inside) | EC_U |
| 13 | Catskill Creek–East End | 182.0 | Tributary (inside) | CC_D |
| 14 | Catskill Creek–First Bridge | 182.2 | Tributary (inside) | CC_U |
| 15 | Normans Kill | 228.5 | Tributary (inside) | NK |
| 16 | Mohawk River at Waterford | 249.4 | Tributary (inside) | MR |
| 17 | Hudson above Mohawk River | 249.6 | Mid-channel | UHR |

[a] STP = Sewage treatment plant outfall; M = mouth of tributary; D = downstream inside of tributary; U = upstream inside of tributary.

### A.2 Sample collection

Prior to every sampling event, new and sealed (never opened) amber glass trace clean bottles were shipped in a cooler to Riverkeeper's offices along the Hudson River. The cooler was loaded on the Riverkeeper patrol boat and samples were collected from the deck of the boat as it moved north from Manhattan to Troy, New York during each sampling month. The formal sampling protocol was as follows: (i) sampler puts on disposable gloves and opens the trace clean bottle with care to not touch the inside or the lip of the bottle; (ii) cap is wrapped in new aluminum foil while sample is collected; (iii) samples are collected upstream from the boat (to avoid collecting stirred up water) using a self-made sampling pole that allows the open bottle to be lowered into the river and rotated; (iv) the uncapped bottle is lowered upside down to approximately 30 cm below the surface of the water; (v) the bottle is tipped to allow water to fill the bottle completely; (vi) the full bottle is retrieved, immediately capped, wrapped in Parafilm, and placed in an ice bath. Once all samples are collected, the filled bottles are placed into the cooler, packed with ice packs, and returned to our laboratory.

**Table A2:** Dates and times of samples collected along the Hudson River Estuary.

| Site Name | May 2016 | July 2016 | Sept. 2016 |
|---|---|---|---|
| Orangetown STP | 2016-05-20 14:30 | 2016-07-13 12:30 | 2016-09-14 13:30 |
| TZ Bridge | 2016-05-20 15:45 | 2016-07-13 13:15 | 2016-09-14 14:30 |
| Pocantico River | 2016-05-20 15:30 | 2016-07-13 14:00 | 2016-09-14 14:15 |
| Cedar Pond Brook | 2016-05-21 12:15 | 2016-07-14 10:45 | 2016-09-15 12:15 |
| Furnace Brook | 2016-05-21 10:45 | 2016-07-14 09:45 | 2016-09-15 10:45 |
| Annesville Creek | 2016-05-21 13:15 | 2016-07-14 12:00 | 2016-09-15 13:30 |
| West Point STP | 2016-05-21 14:45 | 2016-07-14 14:00 | 2016-09-15 14:45 |
| Rondout Creek–Kingston Public Dock | 2016-05-22 10:00 | 2016-07-15 10:15 | 2016-09-16 09:30 |
| Rondout Creek–STP | *sample lost* | 2016-07-15 09:00 | 2016-09-16 09:00 |
| Rondout Creek–Eddyville Anchorage | 2016-05-22 10:30 | 2016-07-15 11:00 | 2016-09-16 10:15 |
| Esopus Creek–Entrance | 2016-05-22 13:30 | 2016-07-15 13:45 | 2016-09-16 15:15 |
| Esopus Creek–West | 2016-05-22 13:15 | 2016-07-15 14:00 | 2016-09-16 15:30 |
| Catskill Creek–East End | 2016-05-22 15:00 | 2016-07-15 16:45 | 2016-09-16 06:45 |
| Catskill Creek–First Bridge | 2016-05-22 15:20 | 2016-07-15 16:15 | 2016-09-16 07:00 |
| Normans Kill | 2016-05-23 14:00 | 2016-07-16 11:45 | 2016-09-17 12:45 |
| Mohawk River at Waterford | 2016-05-23 16:45 | 2016-07-16 14:45 | 2016-09-17 15:15 |
| Hudson above Mohawk River | 2016-05-23 16:30 | 2016-07-16 14:30 | 2016-09-17 15:00 |

| Site Name | May 2017 | June 2017 | July 2017 | Aug. 2017 | Sept. 2017 | Oct. 2017 |
|---|---|---|---|---|---|---|
| Orangetown STP | 2017-05-15 13:45 | 2017-06-12 14:30 | 2017-07-10 13:40 | 2017-08-14 13:45 | 2017-09-11 13:45 | 2017-10-16 13:30 |
| TZ Bridge | *sample lost* | n/a | 2017-07-10 15:15 | n/a | 2017-09-11 14:30 | n/a |
| Pocantico River | 2017-05-15 09:45 | n/a | 2017-07-10 14:30 | n/a | 2017-09-11 14:45 | n/a |
| Cedar Pond Brook | 2017-05-16 11:30 | n/a | 2017-07-11 10:45 | n/a | 2017-09-12 12:15 | n/a |
| Furnace Brook | 2017-05-16 10:30 | n/a | 2017-07-11 09:30 | n/a | 2017-09-12 11:15 | n/a |
| Annesville Creek | 2017-05-16 12:30 | n/a | 2017-07-11 12:45 | n/a | 2017-09-12 13:15 | n/a |
| West Point STP | 2017-05-16 14:00 | 2017-06-13 14:30 | 2017-07-11 14:00 | 2017-08-15 15:00 | 2017-09-12 14:45 | 2017-10-17 15:30 |
| Rondout Creek–Kingston Public Dock | 2017-05-17 08:45 | n/a | 2017-07-12 09:00 | n/a | 2017-09-13 09:00 | n/a |
| Rondout Creek–STP | 2017-05-17 10:00 | 2017-06-14 10:15 | 2017-07-12 10:15 | 2017-08-16 10:00 | 2017-09-13 08:45 | 2017-10-18 09:45 |
| Rondout Creek–Eddyville Anchorage | 2017-05-17 09:15 | 2017-06-14 09:30 | 2017-07-12 09:30 | 2017-08-16 09:15 | 2017-09-13 09:30 | 2017-10-18 10:45 |
| Esopus Creek–Entrance | 2017-05-17 12:15 | n/a | 2017-07-12 12:30 | n/a | 2017-09-13 12:30 | n/a |
| Esopus Creek–West | 2017-05-17 12:30 | 2017-06-14 13:00 | 2017-07-12 12:45 | 2017-08-16 12:45 | 2017-09-13 12:45 | 2017-10-18 13:30 |
| Catskill Creek–East End | 2017-05-18 09:00 | n/a | 2017-07-12 14:45 | n/a | 2017-09-13 14:45 | n/a |
| Catskill Creek–First Bridge | 2017-05-18 09:15 | 2017-06-14 15:15 | 2017-07-12 15:00 | 2017-08-16 15:15 | 2017-09-13 15:00 | 2017-10-18 1530 |
| Normans Kill | 2017-05-18 16:15 | 2017-06-15 13:30 | 2017-07-13 12:45 | 2017-08-17 11:45 | 2017-09-14 12:15 | 2017-10-19 12:15 |
| Mohawk River at Waterford | 2017-05-18 19:00 | 2017-06-15 16:15 | 2017-07-13 15:30 | 2017-08-17 14:15 | 2017-09-14 14:45 | 2017-10-19 15:00 |
| Hudson above Mohawk River | 2017-05-18 18:45 | 2017-06-15 16:00 | 2017-07-13 15:45 | 2017-08-17 14:30 | 2017-09-14 14:30 | 2017-10-19 14:45 |

n/a = not applicable, sample was not planned

We first identified the closest USGS stream gage station to each sample site. Only five of the selected tributaries (RC, EC, NK, MR, and UHR) have stream gages that are currently collecting data. We estimated missing streamflow rates by first calculating the average deviation of the instantaneous streamflow rates from the long-term average streamflow rate of each tributary on each sampling date and then multiplying the long-term average streamflow rate of the missing tributary by the average deviation of the other tributaries on each sampling date. Instantaneous streamflow rates deviated from the long-term averages by approximately the same amount (the standard deviation ranged from ±3% to ±18%), which provides some validation for this approach. **Table A3** describes the long-term and study period average streamflow rates, and the instantaneous streamflow rates of each tributary during each sampling event.

**Table A3:** Stream flowrates ($m^3 \cdot s^{-1}$) for sample sites with USGS stream gage stations.[1]

| **Site:** | **RC** | **EC** | **CC**[a] | **NK** | **MR** | **UHR** |
|---|---|---|---|---|---|---|
| **USGS station** | 01367500 | 01364500 | 01362090 | 01359528 | 01357500 | 01335755 |
| **Long-term average (# Years)** | 17.1 (7) | 14.6 (7) | 18.8 (4) | 4.5 (7) | 181.2 (7) | 199.7 (3) |
| **Study period average** | 6.8 | 5.1 | 7.6 | 2.0 | 101.3 | 179.0 |
| May 2016 | 6.9 | 3.8 | 8.5 | 3.1 | 130.0 | 122.6 |
| July 2016 | 5.0 | 5.0 | 5.2 | 0.8 | 66.0 | 92.3 |
| Sept. 2016 | 3.1 | 1.4 | 2.3 | 0.4 | 35.1 | 58.9 |
| May 2017 | 20.8 | 18.3 | 24.6 | 6.4 | 204.2 | 407.8 |
| June 2017 | 9.8 | 7.6 | 11.7 | 3.4 | 120.6 | 294.5 |
| July 2017 | 4.2 | 3.1 | 5.4 | 1.8 | 238.4 | 337.0 |
| Aug. 2017 | 4.1 | 2.5 | 3.7 | 0.8 | 26.7 | 108.7 |
| Sept. 2017 | 4.1 | 2.2 | 4.2 | 1.2 | 34.8 | 100.0 |
| Oct. 2017 | 3.2 | 1.9 | 2.8 | 0.6 | 55.5 | 89.5 |

[a] Estimated as described in the preceding text

## A.3 Sample preparation and analysis

Stock solutions of all reference standards (Sigma Aldrich) were prepared at 1 $g \cdot L^{-1}$ in either LC-MS-grade methanol (OmniSolv, VWR), nanopure water (EMD Millipore), LC-MS-grade acetonitrile (Fisher Chemical), ethanol (Decon Labs), or dimethyl sulfoxide (Macron Fine Chemicals) and stored at -20°C. A mixture of all reference standards was created in nanopure water at 5 $mg \cdot L^{-1}$ and stored at -20°C. The mixture was diluted with nanopure water to create a calibration curve at 0, 1, 5, 10, 25, 50, 100, 250, 500, 750, and 1000 $ng \cdot L^{-1}$. A list of all target micropollutants is provided in **Table A4**. Similarly, a mixture of isotope-labeled reference standards was created in nanopure water at 5 $mg \cdot L^{-1}$ and stored at -20°C. The mixture was spiked into each calibration standard and sample at a fixed mass of 100 ng prior to sample processing. Reference standard and isotope-labeled reference standard information is provided in **Table A5** and **Table A6**, respectively.

The calibration standards, samples, and method blanks were processed using a mixed-bed solid phase extraction (SPE) method to concentrate the samples.[2] Briefly, 1 L of each calibration standard and each sample was filtered (GF/F, 0.7 μm, Whatman) and pH-adjusted to 6.5 ± 0.1 using 10% formic acid and 1.4 N ammonia (both in nanopure water). Then, the samples were loaded onto custom-packed mixed bed SPE cartridges using a vacuum manifold. The SPE cartridges contained three distinct layers separated by frits (polyethylene, Thermo Fisher Scientific); the top layer consisted of 200 mg of Oasis HLB (Waters); the middle layer consisted of a mixture of 100 mg of Strata X-AW (Phenomenex), 100 mg of Strata X-CW (Phenomenex), and 150 mg of Isolute ENV+ (Biotage); the bottom layer consisted of 200 mg of Envi-Carb (Supelclean). Analytes were eluted from the cartridges using 6 mL of 50:50 ethyl acetate and methanol (v/v) amended with 0.5% ammonia, 3 mL of 50:50 ethyl acetate and methanol (v/v)

amended with 1.7% formic acid, and 2 mL of methanol. The extracts were evaporated to 100 µL under a gentle stream of high-purity nitrogen and reconstituted to 1 mL with nanopure water. Then, the reconstituted extracts were filtered (regenerated cellulose, 0.45 µm, Thermo Fisher Scientific) and stored at -20ºC until analysis.

A previously reported analytical method using high-performance liquid chromatography (HPLC) coupled to high-resolution mass spectrometry was used to quantify the target list of micropollutants.[2,3] We used an HPLC system coupled to a QExactive (quadrupole orbitrap) high resolution mass spectrometer (Thermo Fisher Scientific). The mobile phase consisted of LC-MS-grade water and LC-MS-grade methanol, each amended with 0.1% formic acid (v/v, 98 – 100%, Thermo Fisher Scientific). Samples were injected at 20 µL on an XBridge C-18 analytical column (2.1 x 50 mm, particle size 3.5 µm, Waters) at 25ºC. Mobile phase was pumped at a flowrate of 0.200 mL $min^{-1}$ following a linear gradient (as percent methanol): 10 – 50% for 0 – 4 min, 50 – 95% for 4 – 17 min, and 95% for 17 – 25 min. The column was reconditioned at 10% for 4 minutes prior to the next sample injection. Each sample was analyzed in rapid polarity switching mode with heated electrospray ionization with the following parameters, voltage: +4.0 kV (positive) or -3.3 kV (negative); sheath gas flow: 40 arbitrary units (AU); auxiliary gas flow: 20 AU; capillary temperature: 320ºC; S-lens RF level: 50 AU, auxiliary gas heater temperature: 50ºC. The instrument method acquired full-scan MS data with the following parameters, a mass-to-charge ratio (*m/z*) range of 100 – 800; resolution: 70,000 at 200 *m/z*; automatic gain control (AGC) target: 500,000; maximum injection time: 200 ms. Data dependent $MS^2$ scans were acquired with an inclusion list consisting of the target micropollutants and the following parameters, resolution: 17,500 at 200 *m/z*, AGC target 200,000; maximum injection time: 100 ms; isolation window: 1 *m/z*; underfill ratio: 0.1%, dynamic exclusion: 8 s. Data collection and processing were conducted

using XCalibur v3.1 (Thermo Fisher Scientific). The target micropollutants were quantified using isotope-labeled internal reference standards based on the ratio of the area responses of the target micropollutant to the internal standard and by 1/x weighted linear least-squares regression. Limits of quantification (LOQs) were determined by the lowest linear calibration point with five sticks and the presence of a diagnostic fragment. Method blanks, instrument blanks, and continuing calibration checks were included in the analyses to account for laboratory sources of contamination, sample carryover, and to verify the precision and accuracy of the calibration.

**Table A4:** Micropollutant information.

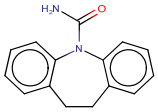
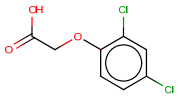
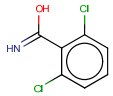
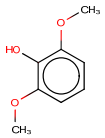
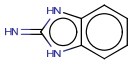
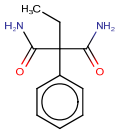
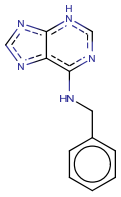
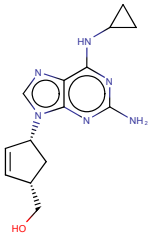
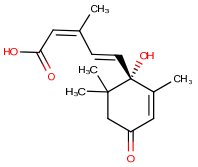
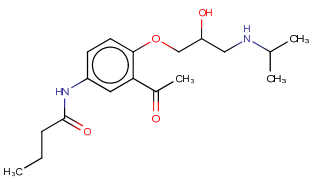
| Micropollutant | Use | Structure | Acidic pKa | Log D (pH = 6) |
|---|---|---|---|---|
| 10,11-dihydrocarbamazepine | Antiepileptic metabolite (carbamazepine) *Wastewater-derived* |  | 15.96 | 2.96 |
| 2,4-D | Herbicide *Agricultural-derived* |  | 2.81 | -0.52 |
| 2,6-dichlorobenzamide | Herbicide degradation product (dichlobenil) *Agricultural-derived* |  | -1.42 | 2.31 |
| 2,6-dimethoxyphenol | Natural component of wood smoke *Wastewater-derived* |  | 9.37 | 1.35 |
| 2-aminobenzimidazole | Fungicide degradation product (carbendazim) *Agricultural-derived* |  | 8.11 | -1.23 |
| 2-ethyl-2-phenyl-malonamide | Pharmaceutical degradation product (primidone) *Wastewater-derived* |  | 15.73 | -2.92 |
| 2-methylisothiazolin-3-one (MI) | Biocide *Agricultural-derived* |  | n/a | 0.23 |
| 6-benzylaminopurine | Plant growth regulator *Agricultural-derived* |  | 3.42 | 1.50 |
| Abacavir | Antiviral *Wastewater-derived* |  | -0.27 | -2.07 |
| Abscisic Acid | Plant hormone and growth regulator *Agricultural-derived* |  | 4.74 | 0.81 |
| Acebutolol | Beta-blocker *Wastewater-derived* |  | 9.57 | -0.25 |

**Table A4** (Continued)

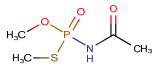
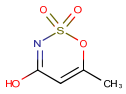
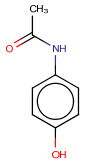
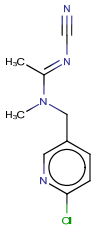
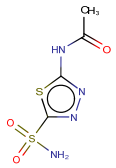
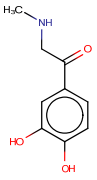
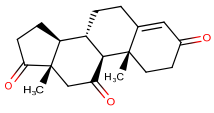
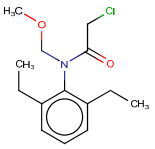
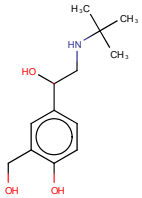
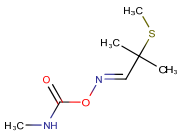
| Micropollutant | Use | Structure | Acidic pKa | Log D (pH = 6) |
|---|---|---|---|---|
| Acephate | Insecticide<br>*Agricultural-derived* |  | 6.54 | 0.03 |
| Acesulfame | Sweetener<br>*Wastewater-derived* |  | 2.00 | -1.97 |
| Acetaminophen | Analgesic<br>*Wastewater-derived* |  | 9.46 | 0.97 |
| Acetamiprid | Neonicotinoid insecticide<br>*Agricultural-derived* |  | -0.27 | 1.11 |
| Acetazolamide | Diuretic<br>*Wastewater-derived* |  | 6.55 | -2.75 |
| Adrenalone | Adrenergic agonist<br>*Wastewater-derived* |  | 7.50 | -1.40 |
| Adrenosterone | Hormone<br>*Wastewater-derived* |  | n/a | 3.01 |
| Alachlor | Herbicide<br>*Agricultural-derived* |  | 16.60 | 3.59 |
| Albuterol | Asthma<br>*Wastewater-derived* |  | 9.40 | -2.21 |
| Aldicarb | Insecticide<br>*Agricultural-derived* |  | 1.63 | -0.16 |

**Table A4** (Continued)

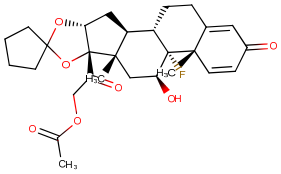
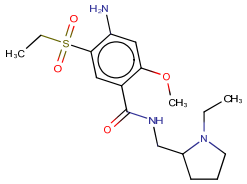
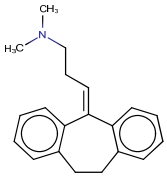
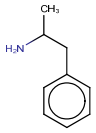
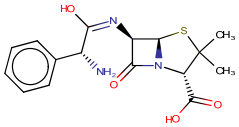
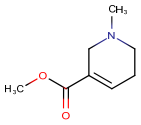
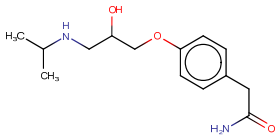
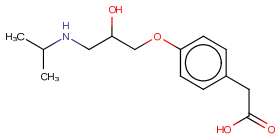
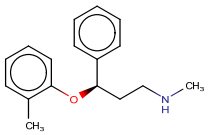
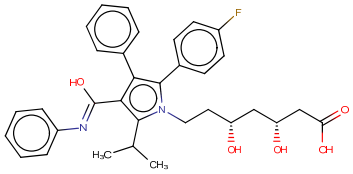
| Micropollutant | Use | Structure | Acidic pKa | Log D (pH = 6) |
|---|---|---|---|---|
| Amcinonide | Glucocorticoid<br>*Wastewater-derived* |  | 13.59 | 3.20 |
| Amisulpride | Antipsychotic<br>*Wastewater-derived* |  | -0.11 | -0.13 |
| Amitriptyline | Antidepressant<br>*Wastewater-derived* |  | 9.76 | 1.50 |
| Amphetamine | Psychostimulant<br>*Wastewater-derived* |  | 10.01 | -1.19 |
| Ampicillin | Antibiotic<br>*Wastewater-derived* |  | 3.24 | -3.49 |
| Arecoline | Alkaloid<br>*Wastewater-derived* |  | 8.23 | -1.56 |
| Atenolol | Beta blocker<br>*Wastewater-derived* |  | 9.67 | -4.50 |
| Atenolol Acid | Pharmaceutical degradation product (atenolol and metoprolol)<br>*Wastewater-derived* |  | 3.54 | -1.24 |
| Atomoxetine | Norepinephrine reuptake inhibitor<br>*Wastewater-derived* |  | 9.80 | 0.67 |
| Atorvastatin | Statins<br>*Wastewater-derived* |  | 4.31 | 3.58 |

**Table A4** (Continued)

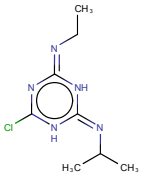
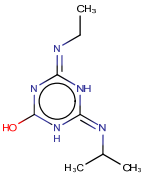
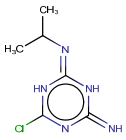
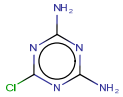
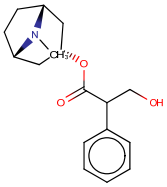
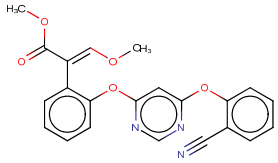
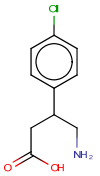
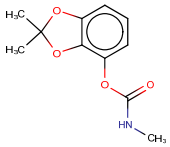
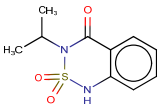
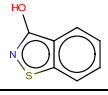
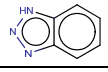
| Micropollutant | Use | Structure | Acidic pKa | Log D (pH = 6) |
|---|---|---|---|---|
| Atrazine | Herbicide<br>*Agricultural-derived* |  | 3.20 | 0.98 |
| Atrazine-2-hydroxy | Herbicide degradation product (atrazine)<br>*Agricultural-derived* |  | 12.48 | -2.84 |
| Atrazine-desethyl | Herbicide degradation product (atrazine)<br>*Agricultural-derived* |  | 3.38 | 0.23 |
| Atrazine-desethyl-desisopropyl | Herbicide degradation product (atrazine)<br>*Agricultural-derived* |  | 3.58 | -0.88 |
| Atropine | Cycloplegic<br>*Wastewater-derived* |  | 9.39 | -1.57 |
| Azoxystrobin | Fungicide<br>*Agricultural-derived* |  | 0.94 | 4.22 |
| Baclofen | GABA receptor agonist<br>*Wastewater-derived* |  | 3.89 | -0.78 |
| Bendiocarb | Insecticide<br>*Agricultural-derived* |  | 14.76 | -0.25 |
| Bentazon | Herbicide<br>*Agricultural-derived* |  | 2.03 | -0.19 |
| Benzisothiazolin-3-one (BIT) | Biocide<br>*Agricultural-derived* |  | 9.50 | 0.66 |
| Benzotriazole | Corrosion inhibitor<br>*Wastewater-derived* |  | 0.22 | 1.30 |

**Table A4** (Continued)

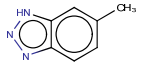
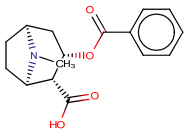
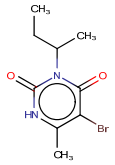
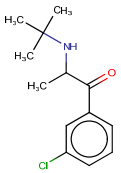
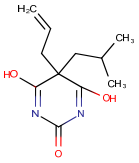
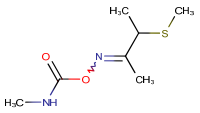
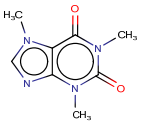
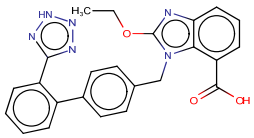
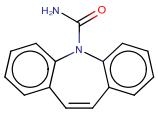
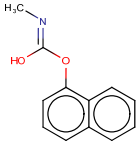
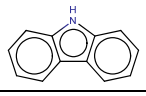
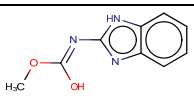
| Micropollutant | Use | Structure | Acidic pKa | Log D (pH = 6) |
|---|---|---|---|---|
| Benzotriazole-methyl-1H | Corrosion inhibitor and antifreeze<br>*Wastewater-derived* |  | 0.45 | 1.81 |
| Benzoylecgonine | Illicit drug metabolite (cocaine)<br>*Wastewater-derived* |  | 3.15 | -0.59 |
| Bromacil | Herbicide<br>*Agricultural-derived* |  | 9.95 | 0.13 |
| Bupropion | Antidepressant<br>*Wastewater-derived* |  | 8.22 | 1.09 |
| Butalbital | Barbiturate<br>*Wastewater-derived* |  | 7.48 | -3.85 |
| Butocarboxim | Insecticide<br>*Agricultural-derived* |  | 1.33 | -0.10 |
| Caffeine | Stimulant<br>*Wastewater-derived* |  | -1.16 | -0.55 |
| Candesartan | Antihypertensive<br>*Wastewater-derived* |  | -1.45 | 2.49 |
| Carbamazepine | Antiepileptic<br>*Wastewater-derived* |  | 15.96 | 2.81 |
| Carbaryl | Insecticide<br>*Agricultural-derived* |  | 14.77 | 1.33 |
| Carbazole | Dye manufacturing<br>*Wastewater-derived* |  | 14.97 | 3.09 |
| Carbendazim | Fungicide<br>*Agricultural-derived* |  | -1.81 | 0.13 |

**Table A4** (Continued)

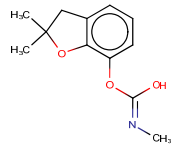
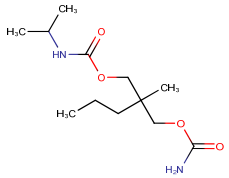
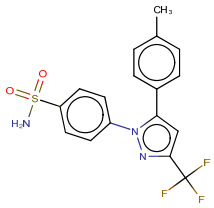
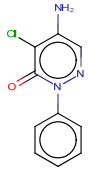
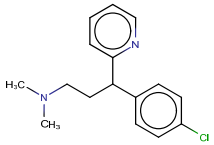
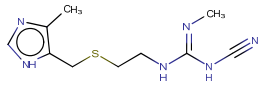
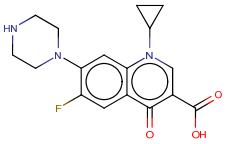
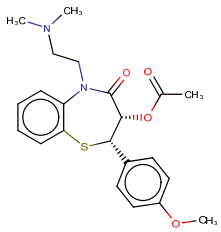
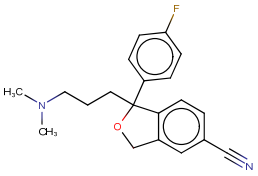
| Micropollutant | Use | Structure | Acidic pKa | Log D (pH = 6) |
|---|---|---|---|---|
| Carbofuran | Insecticide<br>*Agricultural-derived* |  | 14.76 | 0.55 |
| Carisoprodol | Muscle Relax<br>*Wastewater-derived* |  | 15.06 | -0.84 |
| Celecoxib | NSAID<br>*Wastewater-derived* |  | -0.41 | 4.01 |
| Chloridazon | Herbicide<br>*Agricultural-derived* |  | -1.77 | 1.11 |
| Chlorpheniramine | Antihistamine<br>*Wastewater-derived* |  | 3.57 | 0.40 |
| Cimetidine | Histamine H2-receptor antagonist<br>*Wastewater-derived* |  | 4.51 | -0.83 |
| Ciprofloxacin | Antibiotic<br>*Wastewater-derived* |  | -0.22 | -0.95 |
| cis-Diltiazem | Ca channel blocker<br>*Wastewater-derived* |  | 8.18 | 0.57 |
| Citalopram | Antidepressant<br>*Wastewater-derived* |  | 9.78 | 0.44 |

**Table A4** (Continued)

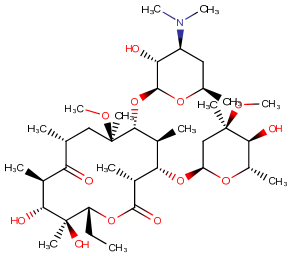
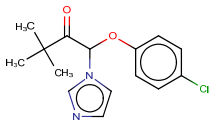
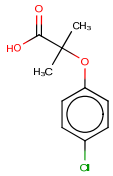
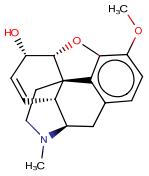
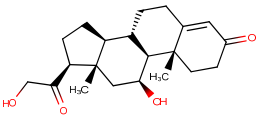
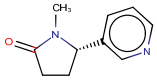
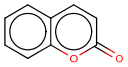
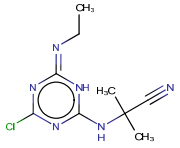
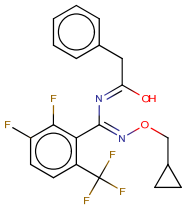
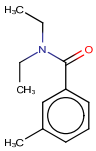
| Micropollutant | Use | Structure | Acidic pKa | Log D (pH = 6) |
|---|---|---|---|---|
| Clarithromycin | Antibiotic<br>*Wastewater-derived* |  | 8.38 | 0.89 |
| Climbazole | Antifungal<br>*Wastewater-derived* |  | 6.49 | 3.98 |
| Clofibric Acid | Herbicide<br>*Agricultural-derived* |  | 3.37 | 0.32 |
| Codeine | Analgesic, antitussive<br>*Wastewater-derived* |  | 9.19 | -1.68 |
| Corticosterone | Hormone<br>*Wastewater-derived* |  | -0.26 | 2.02 |
| Cotinine | Metabolite (nicotine)<br>*Wastewater-derived* |  | -1.83 | 0.19 |
| Coumarin | Anticoagulant<br>*Wastewater-derived* |  | n/a | 1.78 |
| Cyanazine | Herbicide<br>*Agricultural-derived* |  | -0.63 | -0.33 |
| Cyflufenamid | Fungicide<br>*Agricultural-derived* |  | 15.15 | n/a |
| DEET | Insect repellent<br>*Wastewater-derived* |  | -0.95 | 2.50 |

**Table A4** (Continued)

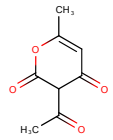
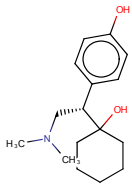
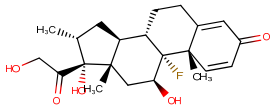
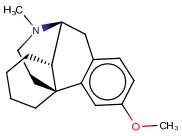
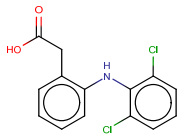
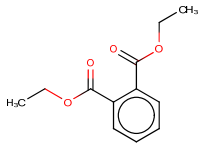
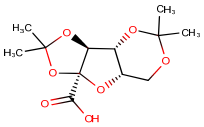
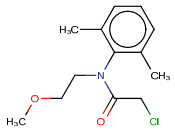
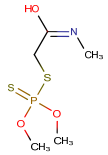
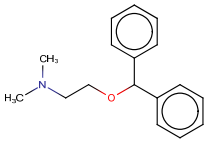
| Micropollutant | Use | Structure | Acidic pKa | Log D (pH = 6) |
|---|---|---|---|---|
| Dehydroacetic Acid | Fungicide<br>*Agricultural-derived* |  | 6.49 | 0.42 |
| Desvenlafaxine | Antidepressant metabolite (venlafaxine)<br>*Wastewater-derived* |  | 8.87 | -0.22 |
| Dexamethasone | Corticosteroid<br>*Wastewater-derived* |  | 12.42 | 1.68 |
| Dextromethorphan | Antitussive<br>*Wastewater-derived* |  | 9.85 | 0.15 |
| Diclofenac | NSAID<br>*Wastewater-derived* |  | 4.00 | 2.26 |
| Diethyl phthalate | Plasticizer<br>*Wastewater-derived* |  | n/a | 2.69 |
| Dikegulac | Growth regulator<br>*Agricultural-derived* |  | 3.05 | n/a |
| Dimethachlor | Herbicide<br>*Agricultural-derived* |  | 16.77 | 2.59 |
| Dimethoate | Insecticide<br>*Agricultural-derived* |  | 15.93 | -1.48 |
| Diphenhydramine | Antihistamine<br>*Wastewater-derived* |  | 8.87 | 0.87 |

**Table A4** (Continued)

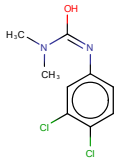
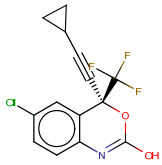
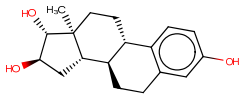
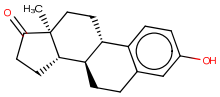
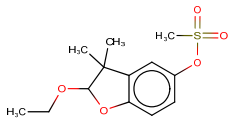
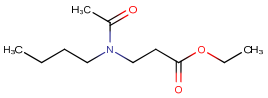
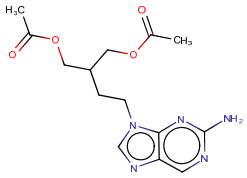
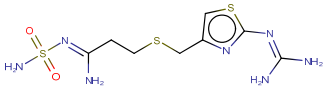
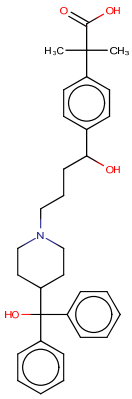
| Micropollutant | Use | Structure | Acidic pKa | Log D (pH = 6) |
|---|---|---|---|---|
| Diuron | Biocide<br>*Agricultural-derived* |  | 13.18 | 2.30 |
| Efavirenz | Antiviral<br>*Wastewater-derived* |  | -1.49 | 1.89 |
| Estriol | Estrogen<br>*Wastewater-derived* |  | 10.33 | 2.67 |
| Estrone | Estrogen<br>*Wastewater-derived* |  | 10.33 | 4.31 |
| Ethofumesate | Herbicide<br>*Agricultural-derived* |  | n/a | 2.34 |
| Ethyl butylacetylaminopropionate | Insect repellent<br>*Wastewater-derived* |  | -1.29 | 0.96 |
| Famciclovir | Antiviral<br>*Wastewater-derived* |  | -1.96 | -1.51 |
| Famotidine | Histamine H2-receptor antagonist<br>*Wastewater-derived* |  | 1.74 | -2.96 |
| Fexofenadine | Antihistamine<br>*Wastewater-derived* |  | 4.04 | 2.94 |

**Table A4** (Continued)

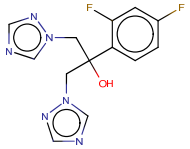
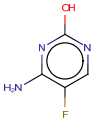
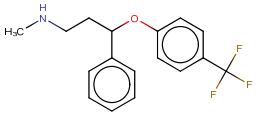
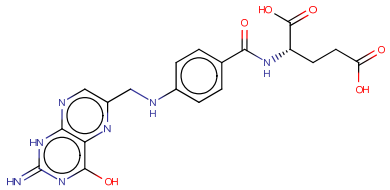
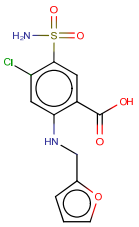
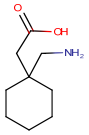
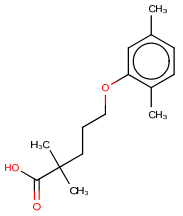
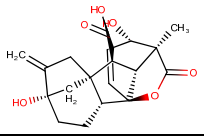
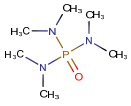
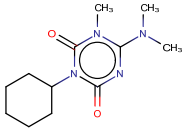
| Micropollutant | Use | Structure | Acidic pKa | Log D (pH = 6) |
|---|---|---|---|---|
| Fluconazole | Antifungal<br>*Wastewater-derived* |  | 1.70 | 0.56 |
| Flucytosine | Antifungal<br>*Wastewater-derived* |  | 8.31 | -0.42 |
| Fluoxetine | Antidepressant<br>*Wastewater-derived* |  | 9.80 | 1.04 |
| Folic Acid | Vitamin<br>*Wastewater-derived* |  | 2.09 | -4.68 |
| Furosemid | Diuretic<br>*Wastewater-derived* |  | -1.52 | 0.00 |
| Gabapentin | Antiepileptic<br>*Wastewater-derived* |  | 4.63 | -1.29 |
| Gemfibrozil | Antilipemic<br>*Wastewater-derived* |  | 4.42 | 2.80 |
| Gibberellic Acid | Plant growth regulator<br>*Agricultural-derived* |  | -0.90 | -1.48 |
| Hexamethylphosphoramide | Phosphoramide<br>*Wastewater-derived* |  | -0.12 | -1.40 |
| Hexazinone | Herbicide<br>*Agricultural-derived* |  | -1.24 | 1.37 |

**Table A4** (Continued)

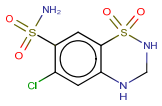
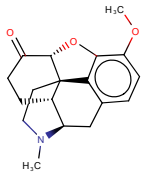
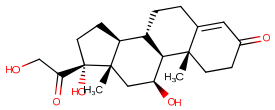
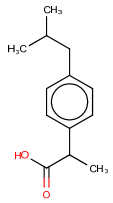
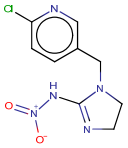
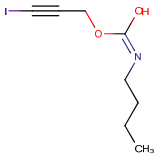
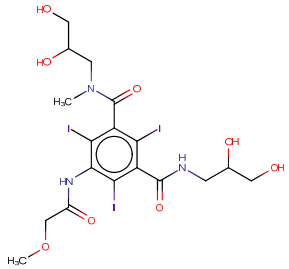
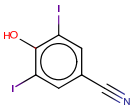
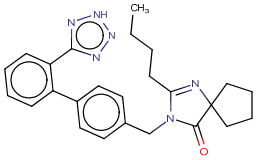
| Micropollutant | Use | Structure | Acidic pKa | Log D (pH = 6) |
|---|---|---|---|---|
| Hydrochlorothiazide | Diuretic<br>*Wastewater-derived* |  | 9.09 | -0.58 |
| Hydrocodone | Analgesic, antitussive<br>*Wastewater-derived* |  | 8.61 | -0.59 |
| Hydrocortisone | Glucocorticoid hormone<br>*Wastewater-derived* |  | 12.59 | 1.28 |
| Ibuprofen | NSAID<br>*Wastewater-derived* |  | 4.85 | 2.67 |
| Imidacloprid | Insecticide<br>*Agricultural-derived* |  | -0.40 | 1.02 |
| Iodocarb | Antifungal<br>*Wastewater-derived* |  | 14.40 | 0.68 |
| Iopromide | Contrast medium<br>*Wastewater-derived* |  | -1.40 | -2.31 |
| Ioxynil | Herbicide<br>*Agricultural-derived* |  | 5.61 | 2.85 |
| Irbesartan | Antihypertensive<br>*Wastewater-derived* |  | -1.47 | 5.13 |

**Table A4** (Continued)

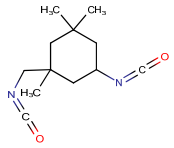
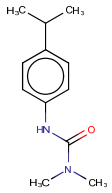
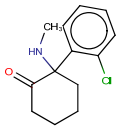
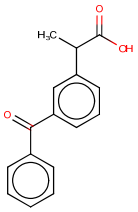
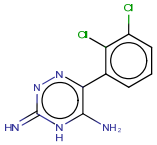
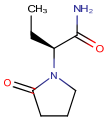
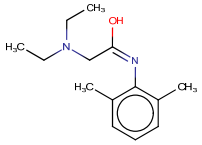
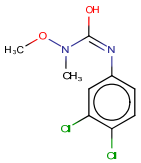
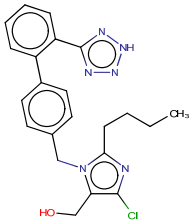
| Micropollutant | Use | Structure | Acidic pKa | Log D (pH = 6) |
|---|---|---|---|---|
| Isophorone Diisocyanate | Isocyanates<br>*Wastewater-derived* |  | n/a | 2.13 |
| Isoproturon | Herbicide<br>*Agricultural-derived* |  | 13.79 | 2.57 |
| Ketamine | Anesthesia<br>*Wastewater-derived* |  | 7.16 | 2.16 |
| Ketoprofen | NSAID<br>*Wastewater-derived* |  | 3.88 | 1.51 |
| Lamotrigine | Anticonvulsant<br>*Wastewater-derived* |  | 5.87 | 1.10 |
| Levetiracetam | Antiepileptic<br>*Wastewater-derived* |  | -1.56 | -2.42 |
| Lidocaine | Topical anesthetic<br>*Wastewater-derived* |  | 7.75 | 0.92 |
| Linuron | Herbicide<br>*Agricultural-derived* |  | 11.94 | 2.30 |
| Losartan | Antihypertensive<br>*Wastewater-derived* |  | -1.45 | 4.72 |

**Table A4** (Continued)

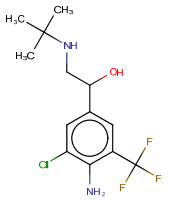
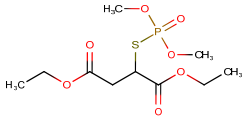
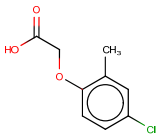
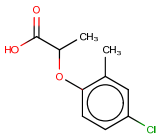
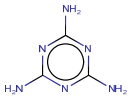
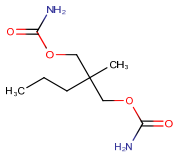
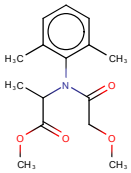
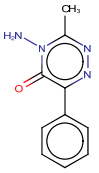
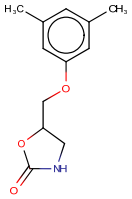
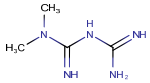
| Micropollutant | Use | Structure | Acidic pKa | Log D (pH = 6) |
|---|---|---|---|---|
| Mabuterol | Anti-asthmatic<br>*Wastewater-derived* |  | 0.90 | -0.48 |
| Malaoxon | Insecticide<br>*Agricultural-derived* |  | n/a | 0.97 |
| MCPA | Herbicide, fungicide<br>*Agricultural-derived* |  | 3.36 | -0.17 |
| Mecoprop | Herbicide<br>*Agricultural-derived* |  | 3.47 | 0.49 |
| Melamine | Insecticide degradation product (cyromazine)<br>*Agricultural-derived* |  | 1.84 | -5.95 |
| Meprobamate | Anxiolytic<br>*Wastewater-derived* |  | 15.17 | -2.79 |
| Metalaxyl | Fungicide<br>*Agricultural-derived* |  | 15.80 | 2.12 |
| Metamitron | Herbicide<br>*Agricultural-derived* |  | 2.78 | 0.44 |
| Metaxalone | Muscle Relax<br>*Wastewater-derived* |  | 13.14 | 1.26 |
| Metformin | Antidiabetic<br>*Wastewater-derived* |  | -1.55 | -5.74 |

**Table A4** (Continued)

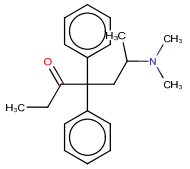
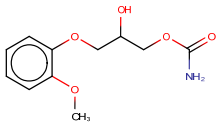
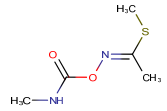
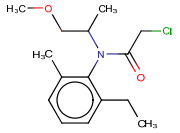
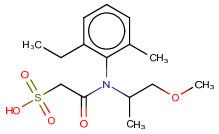
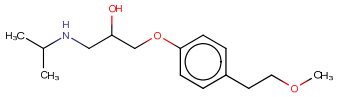
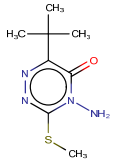
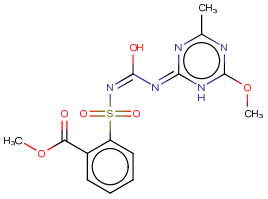
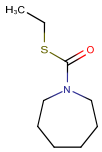
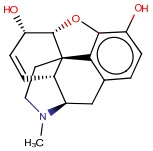
| Micropollutant | Use | Structure | Acidic pKa | Log D (pH = 6) |
|---|---|---|---|---|
| Methadone | Analgesic, narcotic<br>*Wastewater-derived* |  | 9.12 | 2.04 |
| Methocarbamol | Muscle Relax<br>*Wastewater-derived* |  | 13.60 | -1.41 |
| Methomyl | Insecticide<br>*Agricultural-derived* |  | 0.69 | -0.96 |
| Metolachlor | Herbicide<br>*Agricultural-derived* |  | 16.75 | 3.45 |
| Metolachlor-ESA | Herbicide degradation product (metolachlor)<br>*Agricultural-derived* |  | -0.68 | -0.26 |
| Metoprolol | Beta-blocker<br>*Wastewater-derived* |  | 9.67 | -1.34 |
| Metribuzin | Herbicide<br>*Agricultural-derived* |  | 2.46 | 1.96 |
| Metsulfuron-methyl | Herbicide<br>*Agricultural-derived* |  | 0.22 | -0.53 |
| Molinate | Herbicide<br>*Agricultural-derived* |  | n/a | 2.34 |
| Morphine | Narcotic<br>*Wastewater-derived* |  | 9.12 | -1.83 |

**Table A4** (Continued)

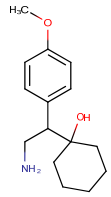
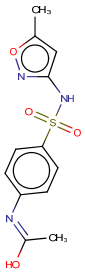
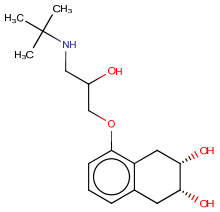
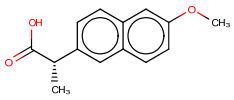
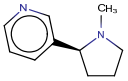
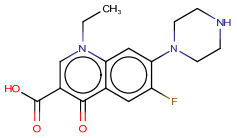
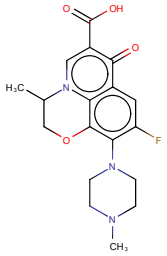
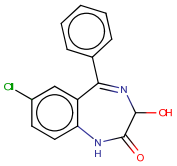
| Micropollutant | Use | Structure | Acidic pKa | Log D (pH = 6) |
|---|---|---|---|---|
| N,N-didesmethyl venlafaxine | Antidepressant metabolite (venlafaxine) *Wastewater-derived* |  | 9.43 | -0.96 |
| N4-acetylsulfamethoxazole | Antibiotic metabolite (sulfamethoxazole) *Wastewater-derived* |  | 0.38 | -0.11 |
| Nadolol | Beta blocker *Wastewater-derived* |  | 9.76 | -2.26 |
| Naproxen | NSAID *Wastewater-derived* |  | 4.19 | 1.18 |
| Nicotine | Stimulant *Wastewater-derived* |  | 2.70 | -1.37 |
| Norfloxacin | Antibiotic *Wastewater-derived* |  | -0.05 | -1.06 |
| Ofloxacin | Antibiotic *Wastewater-derived* |  | 5.45 | 0.65 |
| Oxazepam | Anxiety *Wastewater-derived* |  | -1.47 | 0.15 |

**Table A4** (Continued)

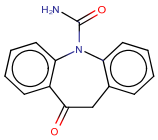
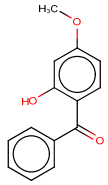
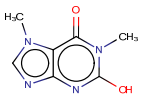
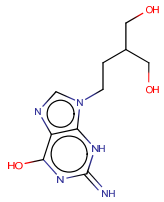
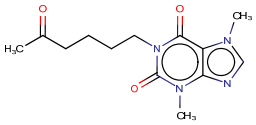
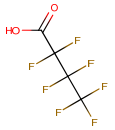
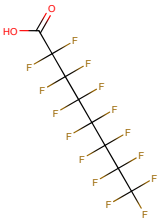
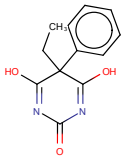
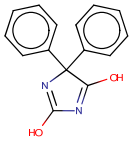
| Micropollutant | Use | Structure | Acidic pKa | Log D (pH = 6) |
|---|---|---|---|---|
| Oxcarbazepine | Anti convulsant<br>*Wastewater-derived* |  | 13.18 | 1.61 |
| Oxybenzone | Sunscreen<br>*Wastewater-derived* |  | 7.07 | 3.59 |
| Paraxanthine | Stimulant metabolite (caffeine)<br>*Wastewater-derived* |  | -1.10 | -2.72 |
| Penciclovir | Antiviral<br>*Wastewater-derived* |  | 2.59 | -3.03 |
| Pentoxyfylline | Cardiovascular<br>*Wastewater-derived* |  | -1.16 | 0.23 |
| Perfluorobutanoic Acid (PFBA) | Perfluorochemical<br>*Wastewater-derived* |  | 1.07 | -1.20 |
| Perfluorooctanoic Acid (PFOA) | Perfluorochemical<br>*Wastewater-derived* |  | -4.20 | 1.58 |
| Phenobarbital | Sedative, anticonvulsant<br>*Wastewater-derived* |  | 7.14 | -4.48 |
| Phenytoin | Antiepileptic<br>*Wastewater-derived* |  | 8.49 | -1.53 |

**Table A4** (Continued)

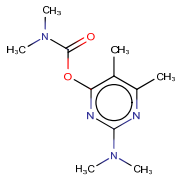
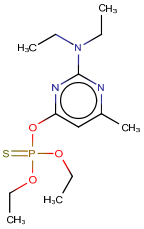
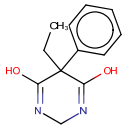
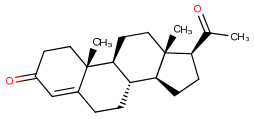
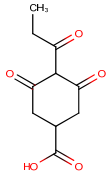
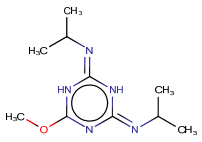
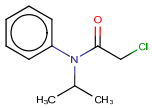
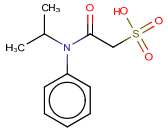
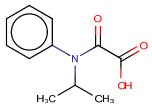
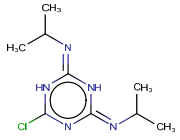
| Micropollutant | Use | Structure | Acidic pKa | Log D (pH = 6) |
|---|---|---|---|---|
| Pirimicarb | Insecticide<br>*Agricultural-derived* |  | 4.99 | 1.76 |
| Pirimiphos-ethyl | Insecticide<br>*Agricultural-derived* |  | 5.09 | 3.63 |
| Primidone | Anti-convulsant<br>*Wastewater-derived* |  | 11.50 | -0.92 |
| Progesterone | Progesterone<br>*Wastewater-derived* |  | n/a | 4.15 |
| Prohexadione | Plant growth regulator<br>*Agricultural-derived* |  | 3.42 | -1.23 |
| Prometon | Herbicide<br>*Agricultural-derived* |  | -1.66 | -0.70 |
| Propachlor | Herbicide<br>*Agricultural-derived* |  | 16.77 | 2.39 |
| Propachlor-ESA | Herbicide metabolite (propachlor)<br>*Agricultural-derived* |  | -0.88 | -1.33 |
| Propachlor-OXA | Herbicide metabolite (propachlor)<br>*Agricultural-derived* |  | 3.03 | -1.06 |
| Propazine | Herbicide<br>*Agricultural-derived* |  | 3.17 | 1.52 |

**Table A4** (Continued)

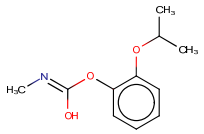
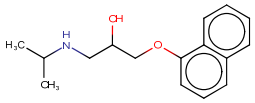
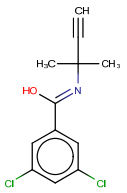
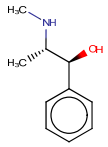
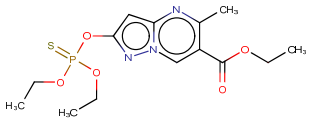
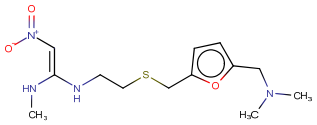
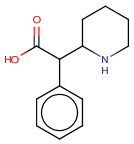
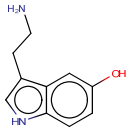
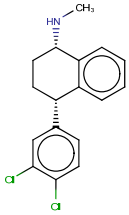
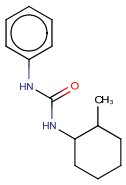
| Micropollutant | Use | Structure | Acidic pKa | Log D (pH = 6) |
|---|---|---|---|---|
| Propoxur | Insecticide<br>*Agricultural-derived* |  | 14.76 | 0.62 |
| Propranolol | Beta Blocker<br>*Wastewater-derived* |  | 9.67 | -0.52 |
| Propyzamide | Herbicide<br>*Agricultural-derived* |  | -1.52 | 3.95 |
| Pseudoephedrine | Decongestant, stimulant<br>*Wastewater-derived* |  | 9.52 | -1.74 |
| Pyrazophos | Fungicide, insecticide<br>*Agricultural-derived* |  | -0.58 | 3.14 |
| Ranitidine | Acid inhibitor<br>*Wastewater-derived* |  | 7.12 | -0.81 |
| Ritalinic Acid | Psychostimulant metabolite (methylphenidate and ethylphenidate)<br>*Wastewater-derived* |  | 3.73 | -0.36 |
| Serotonin | Neurotransmitter<br>*Wastewater-derived* |  | 9.31 | -1.76 |
| Sertraline | Antidepressant<br>*Wastewater-derived* |  | 9.56 | 2.08 |
| Siduron | Herbicide<br>*Agricultural-derived* |  | 13.56 | 2.99 |

**Table A4** (Continued)

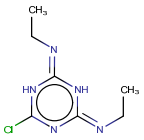
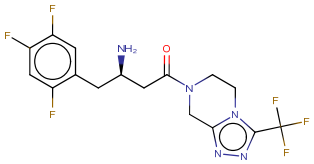
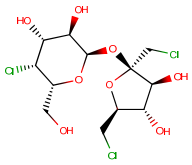
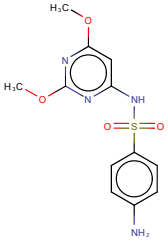
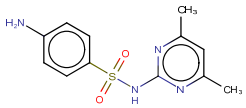
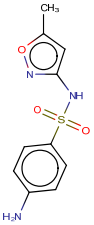
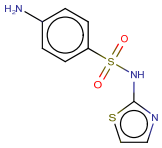
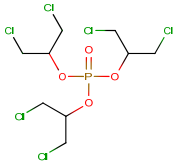
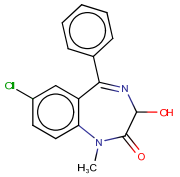
| Micropollutant | Use | Structure | Acidic pKa | Log D (pH = 6) |
|---|---|---|---|---|
| Simazine | Herbicide<br>*Agricultural-derived* |  | 3.23 | 0.58 |
| Sitagliptin | Antidiabetic<br>*Wastewater-derived* |  | 0.66 | -1.33 |
| Sucralose | Sweetener<br>*Wastewater-derived* |  | 11.91 | -0.47 |
| Sulfadimethoxine | Antibiotic<br>*Wastewater-derived* |  | 1.95 | 1.22 |
| Sulfamethazine | Antibiotic<br>*Wastewater-derived* |  | -1.97 | 0.61 |
| Sulfamethoxazole | Antibiotic<br>*Wastewater-derived* |  | 0.25 | 0.60 |
| Sulfathiazole | Antibiotic<br>*Wastewater-derived* |  | 0.35 | 0.93 |
| TDCPP | Flame retardants<br>*Wastewater-derived* |  | n/a | 4.28 |
| Temazepam | Hypnotic<br>*Wastewater-derived* |  | -1.40 | 2.79 |

**Table A4** (Continued)

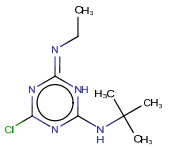
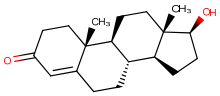
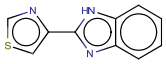
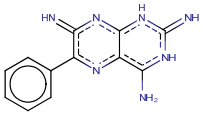
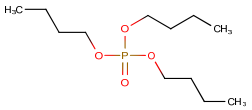
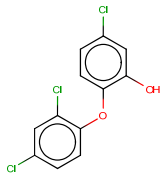
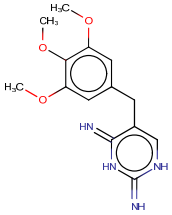
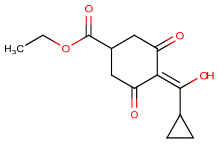
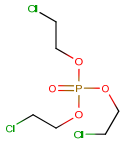
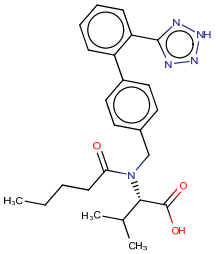
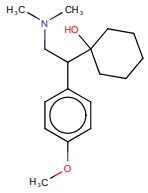
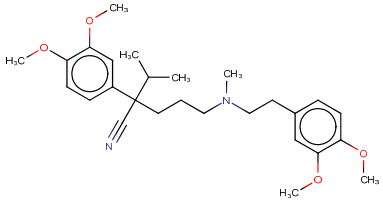
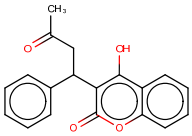
| Micropollutant | Use | Structure | Acidic pKa | Log D (pH = 6) |
|---|---|---|---|---|
| Terbutylazine | Algaecide<br>*Agricultural-derived* |  | 3.18 | -0.35 |
| Testosterone | Androgen<br>*Wastewater-derived* |  | -0.88 | 3.37 |
| Thiabendazole | Fungicide<br>*Wastewater-derived* |  | -0.22 | 2.32 |
| Triamterene | Diuertic<br>*Wastewater-derived* |  | 1.86 | -1.44 |
| Tributyl-phosphate | Plasticizer<br>*Wastewater-derived* |  | n/a | 4.09 |
| Triclosan | Antimicrobial<br>*Wastewater-derived* |  | 7.68 | 4.97 |
| Trimethoprim | Analgesic<br>*Wastewater-derived* |  | -0.90 | -2.35 |
| Trinexapac-ethyl | Herbicide<br>*Agricultural-derived* |  | 4.56 | -1.67 |
| Tris(2-chloroethyl)phosphate | Plasticizer<br>*Wastewater-derived* |  | n/a | 2.11 |
| Valsartan | Angiotensin II antagonist<br>*Wastewater-derived* |  | -1.52 | 3.27 |

**Table A4** (Continued)

| Micropollutant | Use | Structure | Acidic pKa | Log D (pH = 6) |
|---|---|---|---|---|
| Venlafaxine | Antidepressant<br>*Wastewater-derived* |  | 8.91 | -0.07 |
| Verapamil | Ca channel blocker<br>*Wastewater-derived* |  | 9.68 | 1.76 |
| Warfarin | Anticoagulant<br>*Wastewater-derived* |  | 5.56 | 2.17 |

Structures and properties were obtained from ChemAxon (http://www.chemaxon.com).

**Table A5:** Reference standard information.

| Micropollutant | CAS number | Adduct | Exact Mass *(m/z)* | RT *(min)* | Diagnostic Fragment *(m/z)* | LOQ *($ng \cdot L^{-1}$)* | Internal Standard Assignment |
|---|---|---|---|---|---|---|---|
| 10,11-dihydrocarbamazepine | 3564-73-6 | [M+H] | 239.1179 | 7.0 | 194.10 | 1 | Carbamazepine-$^{13}C_6$ |
| 2,4-D | 94-75-7 | [M-H] | 218.9611 | 8.1 | 160.96 | 5 | Mecoprop-$d_3$ |
| 2,6-dichlorobenzamide | 2008-58-4 | [M+H] | 189.9826 | 3.4 | 172.96 | 5 | Dimethoate-$d_3$ |
| 2,6-dimethoxyphenol | 91-10-1 | [M+H] | 155.0708 | 4.7 | 95.05 | 50 | Dimethoate-$d_3$ |
| 2-aminobenzimidazole | 934-32-7 | [M+H] | 134.0713 | 2.7 | 92.05 | 5 | Iopromide-$d_3$ |
| 2-ethyl-2-phenyl-malonamide | 7206-76-0 | [M+H] | 207.1134 | 3.4 | 119.09 | 5 | Iopromide-$d_3$ |
| 2-methylisothiazolin-3-one (MI) | 2682-20-4 | [M+H] | 116.0164 | 1.6 | 100.99 | 1 | 2-methyl-3-isothiazolinone-$d_3$ |
| 6-benzylaminopurine | 1214-39-7 | [M+H] | 226.1093 | 4.6 | 148.06 | 1 | Dimethoate-$d_3$ |
| Abacavir | 136470-78-5 | [M+H] | 287.1620 | 3.4 | 191.10 | 5 | Trimethoprim-$d_9$ |
| Abscisic Acid | 21293-29-8 | [M+H] | 265.1434 | 6.1 | 187.11 | 1 | Tris-2-chloroethyl-phosphate-$d_{12}$ |
| Acebutolol | 37517-30-9 | [M+H] | 337.2122 | 4.4 | 319.20 | 1 | Dimethoate-$d_3$ |
| Acephate | 30560-19-1 | [M+H] | 184.0192 | 2.1 | 142.99 | 5 | Acetaminophen-$d_4$ |
| Acesulfame | 55589-62-3 | [M-H] | 161.9850 | 3.5 | 82.03 | 5 | Mecoprop-$d_3$ |
| Acetaminophen | 103-90-2 | [M+H] | 152.0706 | 2.3 | 110.06 | 5 | Acetaminophen-$d_4$ |
| Acetamiprid | 135410-20-7 | [M+H] | 223.0745 | 4.5 | 126.01 | 5 | Dimethoate-$d_3$ |
| Acetazolamide | 1424-27-7 | [M+H] | 222.9954 | 2.6 | 181.97 | 5 | Metoprolol Acid-$d_5$ |
| Adrenalone | 99-45-6 | [M+H] | 182.0812 | 1.1 | 146.06 | 25 | Metformin-$d_6$ |
| Adrenosterone | 382-45-6 | [M+H] | 301.1803 | 6.6 | 257.15 | 1 | Carbofuran-$d_3$ |
| Alachlor | 15972-60-8 | [M+H] | 270.1255 | 10.3 | 148.11 | 5 | Naproxen-methoxy-$d_3$ |
| Albuterol | 18559-94-9 | [M+H] | 240.1599 | 2.2 | 148.08 | 1 | Metoprolol Acid-$d_5$ |
| Aldicarb | 116-06-3 | [M+Na] | 213.0668 | 5.3 | 89.04 | 50 | Iodocarb-$d_9$ |
| Amcinonide | 51022-69-6 | [M+H] | 503.2445 | 11.9 | 321.15 | 5 | Ibuprofen-$d_3$ |
| Amisulpride | 71675-85-9 | [M+H] | 370.1795 | 3.2 | 242.05 | 1 | Metoprolol Acid-$d_5$ |
| Amitriptyline | 50-48-6 | [M+H] | 278.1908 | 7.8 | 233.13 | 1 | Isoproturon-$d_6$ |
| Amphetamine | 300-62-9 | [M+H] | 136.1126 | 3.3 | 91.05 | 5 | Venlafaxine-$d_6$ |
| Ampicillin | 69-53-4 | [M+H] | 350.1169 | 4.2 | 192.05 | 25 | Pirimicarb-$d_6$ |
| Arecoline | 63-75-2 | [M+H] | 156.1019 | 1.2 | 113.06 | 5 | 2-methyl-3-isothiazolinone-$d_3$ |
| Atenolol | 29122-68-7 | [M+H] | 267.1703 | 2.3 | 190.09 | 1 | Atenolol-$d_7$ |
| Atenolol Acid | 56392-14-4 | [M+H] | 268.1549 | 3.3 | 191.07 | 5 | Metoprolol Acid-$d_5$ |
| Atomoxetine | 83015-26-3 | [M+H] | 256.1696 | 7.1 | 224.08 | 1 | Atrazine-$d_5$ |
| Atorvastatin | 134523-00-5 | [M+H] | 559.2608 | 11.0 | 440.22 | 100 | Ibuprofen-$d_3$ |
| Atrazine | 1912-24-9 | [M+H] | 216.1010 | 7.3 | 174.05 | 1 | Atrazine-$d_5$ |
| Atrazine-2-hydroxy | 2163-68-0 | [M+H] | 198.1349 | 3.8 | 156.09 | 1 | Atrazine-$d_5$ |
| Atrazine-desethyl | 6190-65-4 | [M+H] | 188.0697 | 4.8 | 146.02 | 1 | Atrazine-$d_5$ |
| Atrazine-desethyl-desisopropyl | 3397-62-4 | [M+H] | 146.0233 | 1.7 | 104.00 | 25 | Atrazine-$d_5$ |
| Atropine | 51-55-8 | [M+H] | 290.1751 | 3.7 | 260.16 | 1 | Thiabendazole-$d_4$ |

**Table A5** (Continued)

| Micropollutant | CAS number | Adduct | Exact Mass *(m/z)* | RT *(min)* | Diagnostic Fragment *(m/z)* | LOQ *($ng \cdot L^{-1}$)* | Internal Standard Assignment |
|---|---|---|---|---|---|---|---|
| 10,11-dihydrocarbamazepine | 3564-73-6 | [M+H] | 239.1179 | 7.0 | 194.10 | 1 | Carbamazepine-$^{13}C_6$ |
| 2,4-D | 94-75-7 | [M-H] | 218.9611 | 8.1 | 160.96 | 5 | Mecoprop-$d_3$ |
| 2,6-dichlorobenzamide | 2008-58-4 | [M+H] | 189.9826 | 3.4 | 172.96 | 5 | Dimethoate-$d_3$ |
| 2,6-dimethoxyphenol | 91-10-1 | [M+H] | 155.0708 | 4.7 | 95.05 | 50 | Dimethoate-$d_3$ |
| 2-aminobenzimidazole | 934-32-7 | [M+H] | 134.0713 | 2.7 | 92.05 | 5 | Iopromide-$d_3$ |
| 2-ethyl-2-phenyl-malonamide | 7206-76-0 | [M+H] | 207.1134 | 3.4 | 119.09 | 5 | Iopromide-$d_3$ |
| 2-methylisothiazolin-3-one (MI) | 2682-20-4 | [M+H] | 116.0164 | 1.6 | 100.99 | 1 | 2-methyl-3-isothiazolinone-$d_3$ |
| 6-benzylaminopurine | 1214-39-7 | [M+H] | 226.1093 | 4.6 | 148.06 | 1 | Dimethoate-$d_3$ |
| Abacavir | 136470-78-5 | [M+H] | 287.1620 | 3.4 | 191.10 | 5 | Trimethoprim-$d_9$ |
| Abscisic Acid | 21293-29-8 | [M+H] | 265.1434 | 6.1 | 187.11 | 1 | Tris-2-chloroethyl-phosphate-$d_{12}$ |
| Acebutolol | 37517-30-9 | [M+H] | 337.2122 | 4.4 | 319.20 | 1 | Dimethoate-$d_3$ |
| Acephate | 30560-19-1 | [M+H] | 184.0192 | 2.1 | 142.99 | 5 | Acetaminophen-$d_4$ |
| Acesulfame | 55589-62-3 | [M-H] | 161.9850 | 3.5 | 82.03 | 5 | Mecoprop-$d_3$ |
| Acetaminophen | 103-90-2 | [M+H] | 152.0706 | 2.3 | 110.06 | 5 | Acetaminophen-$d_4$ |
| Acetamiprid | 135410-20-7 | [M+H] | 223.0745 | 4.5 | 126.01 | 5 | Dimethoate-$d_3$ |
| Acetazolamide | 1424-27-7 | [M+H] | 222.9954 | 2.6 | 181.97 | 5 | Metoprolol Acid-$d_5$ |
| Adrenalone | 99-45-6 | [M+H] | 182.0812 | 1.1 | 146.06 | 25 | Metformin-$d_6$ |
| Adrenosterone | 382-45-6 | [M+H] | 301.1803 | 6.6 | 257.15 | 1 | Carbofuran-$d_3$ |
| Alachlor | 15972-60-8 | [M+H] | 270.1255 | 10.3 | 148.11 | 5 | Naproxen-methoxy-$d_3$ |
| Albuterol | 18559-94-9 | [M+H] | 240.1599 | 2.2 | 148.08 | 1 | Metoprolol Acid-$d_5$ |
| Aldicarb | 116-06-3 | [M+Na] | 213.0668 | 5.3 | 89.04 | 50 | Iodocarb-$d_9$ |
| Amcinonide | 51022-69-6 | [M+H] | 503.2445 | 11.9 | 321.15 | 5 | Ibuprofen-$d_3$ |
| Amisulpride | 71675-85-9 | [M+H] | 370.1795 | 3.2 | 242.05 | 1 | Metoprolol Acid-$d_5$ |
| Amitriptyline | 50-48-6 | [M+H] | 278.1908 | 7.8 | 233.13 | 1 | Isoproturon-$d_6$ |
| Amphetamine | 300-62-9 | [M+H] | 136.1126 | 3.3 | 91.05 | 5 | Venlafaxine-$d_6$ |
| Ampicillin | 69-53-4 | [M+H] | 350.1169 | 4.2 | 192.05 | 25 | Pirimicarb-$d_6$ |
| Arecoline | 63-75-2 | [M+H] | 156.1019 | 1.2 | 113.06 | 5 | 2-methyl-3-isothiazolinone-$d_3$ |
| Atenolol | 29122-68-7 | [M+H] | 267.1703 | 2.3 | 190.09 | 1 | Atenolol-$d_7$ |
| Atenolol Acid | 56392-14-4 | [M+H] | 268.1549 | 3.3 | 191.07 | 5 | Metoprolol Acid-$d_5$ |
| Atomoxetine | 83015-26-3 | [M+H] | 256.1696 | 7.1 | 224.08 | 1 | Atrazine-$d_5$ |
| Atorvastatin | 134523-00-5 | [M+H] | 559.2608 | 11.0 | 440.22 | 100 | Ibuprofen-$d_3$ |
| Atrazine | 1912-24-9 | [M+H] | 216.1010 | 7.3 | 174.05 | 1 | Atrazine-$d_5$ |
| Atrazine-2-hydroxy | 2163-68-0 | [M+H] | 198.1349 | 3.8 | 156.09 | 1 | Atrazine-$d_5$ |
| Atrazine-desethyl | 6190-65-4 | [M+H] | 188.0697 | 4.8 | 146.02 | 1 | Atrazine-$d_5$ |

**Table A5** (Continued)

| Micropollutant | CAS number | Adduct | Exact Mass | RT | Diagnostic Fragment | LOQ | Internal Standard Assignment |
|---|---|---|---|---|---|---|---|
| Atrazine-desethyl-desisopropyl | 3397-62-4 | [M+H] | 146.0233 | 1.7 | 104.00 | 25 | Atrazine-$d_5$ |
| Atropine | 51-55-8 | [M+H] | 290.1751 | 3.7 | 260.16 | 1 | Thiabendazole-$d_4$ |
| Azoxystrobin | 131860-33-8 | [M+H] | 404.1241 | 8.7 | 372.10 | 1 | Azoxystrobin-$d_4$ |
| Baclofen | 1134-47-0 | [M+H] | 214.0635 | 3.5 | 151.03 | 5 | Thiabendazole-$d_4$ |
| Bendiocarb | 22781-23-3 | [M+H] | 224.0917 | 6.2 | 167.07 | 5 | Tris-2-chloroethyl-phosphate-$d_{12}$ |
| Bentazon | 25057-89-0 | [M-H] | 239.0485 | 6.6 | 175.09 | 5 | Mecoprop-$d_3$ |
| Benzisothiazolin-3-one (BIT) | 2634-33-5 | [M+H] | 152.0164 | 4.4 | 109.01 | 5 | Carbofuran-$d_3$ |
| Benzotriazole | 95-14-7 | [M+H] | 120.0556 | 3.9 | 92.05 | 1 | Thiabendazole-$d_4$ |
| Benzotriazole-methyl-1H | 136-85-6 | [M+H] | 134.0713 | 5.1 | 106.07 | 1 | Tris-2-chloroethyl-phosphate-$d_{12}$ |
| Benzoylecgonine | 519-09-5 | [M+H] | 290.1387 | 4.0 | 168.10 | 1 | Thiabendazole-$d_4$ |
| Bromacil | 314-40-9 | [M+H] | 261.0233 | 6.1 | 204.96 | 25 | Carbofuran-$d_3$ |
| Bupropion | 34841-39-9 | [M+H] | 240.1155 | 5.1 | 166.04 | 25 | Citalopram-$d_6$ |
| Butalbital | 77-26-9 | [M-H] | 223.1083 | 6.1 | 136.37 | 1 | Mecoprop-$d_3$ |
| Butocarboxim | 34681-10-2 | [M+H] | 191.0849 | 5.4 | n/a | 250 | Venlafaxine-$d_6$ |
| Caffeine | 58-08-2 | [M+H] | 195.0876 | 3.4 | 138.07 | 25 | Caffeine-$^{13}C_3$ |
| Candesartan | 139481-59-7 | [M+H] | 441.1670 | 8.2 | 263.13 | 1 | Iodocarb-$d_9$ |
| Carbamazepine | 298-46-4 | [M+H] | 237.1022 | 6.6 | 194.10 | 1 | Carbamazepine-$^{13}C_6$ |
| Carbaryl | 63-25-2 | [M+H] | 202.0873 | 6.6 | 145.06 | 1 | Carbaryl-$d_7$ |
| Carbazole | 86-74-8 | [M+H] | 168.0813 | 9.6 | 106.96 | 100 | Oxybenzone-$d_3$ |
| Carbendazim | 10605-21-7 | [M+H] | 192.0768 | 3.1 | 160.05 | 1 | Metoprolol Acid-$d_5$ |
| Carbofuran | 1563-66-2 | [M+H] | 222.1135 | 6.1 | 165.09 | 25 | Carbofuran-$d_3$ |
| Carisoprodol | 78-44-4 | [M+H] | 261.1814 | 7.6 | 158.12 | 5 | Isoproturon-$d_6$ |
| Celecoxib | 169590-42-5 | [M+H] | 382.0837 | 10.9 | 362.08 | 5 | Celecoxib-$d_4$ |
| Chloridazon | 1698-60-8 | [M+H] | 222.0428 | 4.5 | 128.99 | 1 | Acetaminophen-$d_4$ |
| Chlorpheniramine | 132-22-9 | [M+H] | 275.1315 | 4.9 | 230.07 | 1 | Venlafaxine-$d_6$ |
| Cimetidine | 51481-61-9 | [M+H] | 253.1235 | 2.3 | 159.07 | 5 | Cimetidine-$d_3$ |
| Ciprofloxacin | 85721-33-1 | [M+H] | 332.1405 | 3.8 | 245.11 | 5 | Imidacloprid-$d_4$ |
| cis-Diltiazem | 42399-41-7 | [M+H] | 415.1691 | 6.4 | 310.17 | 1 | Diltiazem-$d_4$ |
| Citalopram | 59729-33-8 | [M+H] | 325.1716 | 5.9 | 262.10 | 1 | Citalopram-$d_6$ |
| Clarithromycin | 81103-11-9 | [M+H] | 748.4842 | 8.8 | 590.39 | 5 | Azoxystrobin-$d_4$ |
| Climbazole | 38083-17-9 | [M+H] | 293.1057 | 7.0 | 197.07 | 1 | Atrazine-$d_5$ |
| Clofibric Acid | 882-09-7 | [M-H] | 213.0313 | 8.9 | 126.99 | 5 | Mecoprop-$d_3$ |
| Codeine | 76-57-3 | [M+H] | 300.1599 | 2.5 | 215.11 | 1 | Metoprolol Acid-$d_5$ |
| Corticosterone | 50-22-6 | [M+H] | 347.2217 | 7.9 | 329.21 | 5 | Azoxystrobin-$d_4$ |
| Cotinine | 486-56-6 | [M+H] | 177.1028 | 1.2 | 146.06 | 1 | Allopurinol-$d_2$ |
| Coumarin | 91-64-5 | [M+H] | 147.0441 | 4.9 | 91.05 | 25 | Citalopram-$d_6$ |

**Table A5** (Continued)

| Micropollutant | CAS number | Adduct | Exact Mass | RT | Diagnostic Fragment | LOQ | Internal Standard Assignment |
|---|---|---|---|---|---|---|---|
| Cyanazine | 21725-46-2 | [M+H] | 241.0963 | 5.7 | 214.09 | 25 | Dextromethorphan-$d_3$ |
| Cyflufenamid | 180409-60-3 | [M+H] | 413.1288 | 12.2 | 241.04 | 5 | Ibuprofen-$d_3$ |
| DEET | 134-62-3 | [M+H] | 192.1383 | 7.3 | 119.05 | 1 | Carbamazepine-$^{13}C_6$ |
| Dehydroacetic Acid | 520-45-6 | [M+H] | 169.0495 | 5.0 | 85.03 | 1 | Venlafaxine-$d_6$ |
| Desvenlafaxine | 93413-62-8 | [M+H] | 264.1958 | 4.2 | 201.13 | 1 | Venlafaxine-$d_6$ |
| Dexamethasone | 50-02-2 | [M+H] | 393.2077 | 8.0 | 237.13 | 5 | Carbofuran-$d_3$ |
| Dextromethorphan | 125-71-3 | [M+H] | 272.2014 | 5.8 | 147.08 | 1 | Dextromethorphan-$d_3$ |
| Diclofenac | 15307-86-5 | [M+H] | 296.0239 | 11.5 | 215.05 | 25 | Diclofenac-$^{13}C_6$ |
| Diethyl phthalate | 84-66-2 | [M+H] | 223.0965 | 7.7 | 149.02 | 25 | Isoproturon-$d_6$ |
| Dikegulac | 52508-35-7 | [M+H] | 275.1125 | 5.2 | 230.07 | 500 | Dimethoate-$d_6$ |
| Dimethachlor | 50563-36-5 | [M+H] | 256.1099 | 7.9 | 224.08 | 1 | Isoproturon-$d_6$ |
| Dimethoate | 60-51-5 | [M+H] | 230.0080 | 4.5 | 142.99 | 1 | Dimethoate-$d_6$ |
| Diphenhydramine | 58-73-1 | [M+H] | 256.1701 | 6.0 | 167.09 | 1 | Citalopram-$d_6$ |
| Diuron | 330-54-1 | [M+H] | 233.0243 | 7.9 | 159.97 | 5 | Isoproturon-$d_6$ |
| Efavirenz | 154598-52-4 | [M+H] | 316.0352 | 11.5 | 244.01 | 25 | Efavirenz-$d_6$ |
| Estriol | 50-27-1 | [M+Na] | 311.1618 | 5.9 | 226.89 | 25 | Fexofenadine-$d_6$ |
| Estrone | 53-16-7 | [M+H] | 271.1698 | 5.8 | 157.06 | 25 | Carbofuran-$d_3$ |
| Ethofumesate | 26225-79-6 | [M+H] | 287.0947 | 8.9 | 161.06 | 5 | Azoxystrobin-$d_4$ |
| Ethyl butylacetylaminopropionate | 52304-36-6 | [M+H] | 216.1599 | 6.6 | 86.10 | 5 | Carbamazepine-$^{13}C_6$ |
| Famciclovir | 104227-87-4 | [M+H] | 322.1510 | 3.6 | 202.11 | 5 | Thiabendazole-$d_4$ |
| Famotidine | 76824-35-6 | [M+H] | 338.0527 | 2.3 | 189.03 | 25 | Cimetidine-$d_3$ |
| Fexofenadine | 83799-24-0 | [M+H] | 502.2957 | 7.4 | 484.28 | 1 | Fexofenadine-$d_6$ |
| Fluconazole | 86386-73-4 | [M+H] | 307.1113 | 4.5 | 220.07 | 1 | Dimethoate-$d_6$ |
| Flucytosine | 2022-85-7 | [M+H] | 130.0411 | 1.1 | 71.06 | 250 | 2-methyl-3-isothiazolinone-$d_3$ |
| Fluoxetine | 54910-89-3 | [M+H] | 310.1413 | 8.4 | 265.16 | 5 | Fluoxetine-$d_5$ |
| Folic Acid | 59-30-3 | [M+H] | 442.1470 | 3.3 | 295.09 | 25 | Metoprolol Acid-$d_5$ |
| Furosemid | 54-31-9 | [M-H] | 328.9999 | 5.7 | 204.98 | 25 | Sucralose-$d_6$ |
| Gabapentin | 60142-96-3 | [M+H] | 172.1332 | 3.2 | 137.10 | 1 | Metoprolol Acid-$d_5$ |
| Gemfibrozil | 25812-30-0 | [M+H] | 251.1641 | 13.5 | 129.09 | 5 | Gemfibrozil-$d_6$ |
| Gibberellic Acid | 77-06-5 | [M-H] | 345.1333 | 4.6 | 221.13 | 25 | Sucralose-$d_6$ |
| Hexamethylphosphoramide | 680-31-9 | [M+H] | 180.1260 | 4.7 | 135.07 | 5 | Dimethoate-$d_6$ |
| Hexazinone | 51235-04-2 | [M+H] | 253.1659 | 6.1 | 171.09 | 1 | Carbofuran-$d_3$ |
| Hydrochlorothiazide | 58-93-5 | [M+H] | 297.9712 | 2.3 | 232.98 | 50 | Acetaminophen-$d_4$ |
| Hydrocodone | 125-29-1 | [M+H] | 300.1599 | 2.8 | 199.08 | 5 | Morphine-$d_3$ |
| Hydrocortisone | 50-23-7 | [M+H] | 363.2171 | 7.0 | 327.19 | 5 | Carbaryl-$d_7$ |
| Ibuprofen | 15687-27-1 | [M+Na] | 229.1199 | 11.9 | 181.99 | 25 | Ibuprofen-$d_3$ |

**Table A5** (Continued)

| Micropollutant | CAS number | Adduct | Exact Mass | RT | Diagnostic Fragment | LOQ | Internal Standard Assignment |
|---|---|---|---|---|---|---|---|
| Imidacloprid | 138261-41-3 | [M+H] | 256.0601 | 4.0 | 175.10 | 5 | Imidacloprid-$d_4$ |
| Iodocarb | 55406-53-6 | [M+H] | 281.9985 | 7.4 | 164.92 | 100 | Iodocarb-$d_9$ |
| Iopromide | 73334-07-3 | [M+H] | 791.8770 | 2.5 | 572.90 | 50 | Iopromide-$d_3$ |
| Ioxynil | 1689-83-4 | [M-H] | 369.8221 | 8.3 | 230.92 | 1 | Sucralose-$d_6$ |
| Irbesartan | 138402-11-6 | [M+H] | 429.2397 | 7.9 | 386.22 | 100 | Isoproturon-$d_6$ |
| Isophorone Diisocyanate | 4098-71-9 | [M+K] | 261.0995 | 7.0 | 204.96 | 5 | Isoproturon-$d_6$ |
| Isoproturon | 34123-59-6 | [M+H] | 207.1492 | 7.7 | 134.10 | 1 | Isoproturon-$d_6$ |
| Ketamine | 6740-88-1 | [M+H] | 238.0998 | 3.9 | 179.06 | 1 | Thiabendazole-$d_4$ |
| Ketoprofen | 22071-15-4 | [M+H] | 255.1015 | 8.5 | 209.10 | 1 | Carbofuran-$d_3$ |
| Lamotrigine | 84057-84-1 | [M+H] | 256.0151 | 4.7 | 210.98 | 1 | Dimethoate-$d_6$ |
| Levetiracetam | 102767-28-2 | [M+H] | 171.1123 | 2.8 | 126.09 | 1 | Iopromide-$d_3$ |
| Lidocaine | 137-58-6 | [M+H] | 235.1810 | 3.7 | 86.10 | 1 | Thiabendazole-$d_4$ |
| Linuron | 330-55-2 | [M+H] | 249.0192 | 8.9 | 159.97 | 5 | Naproxen-methoxy-$d_3$ |
| Losartan | 114798-26-4 | [M+H] | 423.1695 | 7.9 | 235.10 | 1 | Isoproturon-$d_6$ |
| Mabuterol | 56341-08-3 | [M+H] | 311.1138 | 4.7 | 237.04 | 1 | Dimethoate-$d_6$ |
| Malaoxon | 1634-78-2 | [M+H] | 315.0667 | 6.2 | 255.07 | 100 | Carbamazepine-$^{13}C_6$ |
| MCPA | 94-74-6 | [M-H] | 199.0157 | 8.5 | 141.01 | 5 | Mecoprop-$d_3$ |
| Mecoprop | 93-65-2 | [M-H] | 213.0313 | 9.7 | 141.01 | 5 | Mecoprop-$d_3$ |
| Melamine | 108-78-1 | [M+H] | 127.0727 | 1.1 | 85.05 | 250 | Thiabendazole-$d_4$ |
| Meprobamate | 57-53-4 | [M+H] | 219.1345 | 5.4 | 203.14 | 5 | Venlafaxine-$d_6$ |
| Metalaxyl | 57837-19-1 | [M+H] | 280.1543 | 7.6 | 160.11 | 1 | Isoproturon-$d_6$ |
| Metamitron | 41394-05-2 | [M+H] | 203.0938 | 4.3 | 175.10 | 5 | Imidacloprid-$d_4$ |
| Metaxalone | 1665-48-1 | [M+H] | 222.1135 | 7.1 | 133.10 | 1 | Atrazine-$d_5$ |
| Metformin | 657-24-9 | [M+H] | 130.1092 | 1.1 | 86.00 | 25 | Metformin-$d_6$ |
| Methadone | 76-99-3 | [M+H] | 310.2171 | 7.5 | 265.16 | 1 | Atrazine-$d_5$ |
| Methocarbamol | 532-03-6 | [M+H] | 242.1028 | 4.7 | 163.08 | 5 | Dimethoate-$d_6$ |
| Methomyl | 16752-77-5 | [M+H] | 163.0541 | 3.2 | 102.97 | 5 | Thiabendazole-$d_4$ |
| Metolachlor | 51218-45-2 | [M+H] | 284.1412 | 10.5 | 252.12 | 1 | Naproxen-methoxy-$d_3$ |
| Metolachlor-ESA | 171118-09-5 | [M+H] | 330.1370 | 6.6 | 298.11 | 1 | Carbamazepine-$^{13}C_6$ |
| Metoprolol | 37350-58-6 | [M+H] | 268.1907 | 4.4 | 159.08 | 1 | Dimethoate-$d_6$ |
| Metribuzin | 21087-64-9 | [M+H] | 215.0961 | 6.0 | 187.10 | 1 | Carbamazepine-$^{13}C_6$ |
| Metsulfuron-methyl | 74223-64-6 | [M+H] | 382.0816 | 6.1 | 340.12 | 25 | Carbofuran-$d_3$ |
| Molinate | 2212-67-1 | [M+H] | 188.1109 | 9.5 | 126.09 | 5 | Azoxystrobin-$d_4$ |
| Morphine | 57-27-2 | [M+H] | 286.1437 | 1.6 | 201.09 | 5 | Morphine-$d_3$ |
| N,N-didesmethyl venlafaxine | 93413-77-5 | [M+H] | 250.1802 | 5.5 | 147.08 | 1 | Venlafaxine-$d_6$ |
| N4-acetylsulfamethoxazole | 21312-10-7 | [M+H] | 296.0700 | 4.8 | 198.02 | 1 | Dimethoate-$d_6$ |

**Table A5** (Continued)

| Micropollutant | CAS number | Adduct | Exact Mass | RT | Diagnostic Fragment | LOQ | Internal Standard Assignment |
|---|---|---|---|---|---|---|---|
| Nadolol | 42200-33-9 | [M+H] | 310.2013 | 3.6 | 254.14 | 1 | Thiabendazole-$d_4$ |
| Naproxen | 22204-53-1 | [M+H] | 231.1015 | 8.9 | 185.10 | 1 | Naproxen-methoxy-$d_3$ |
| Nicotine | 54-11-5 | [M+H] | 163.1235 | 1.1 | 102.97 | 5 | Metformin-$d_6$ |
| Norfloxacin | 70458-96-7 | [M+H] | 320.1405 | 3.7 | 276.15 | 5 | Metoprolol Acid-$d_5$ |
| Ofloxacin | 82419-36-1 | [M+H] | 362.1511 | 3.5 | 318.16 | 5 | Metoprolol Acid-$d_5$ |
| Oxazepam | 604-75-1 | [M+H] | 287.0558 | 7.8 | 241.05 | 5 | Azoxystrobin-$d_4$ |
| Oxcarbazepine | 28721-07-5 | [M+H] | 253.0977 | 5.7 | 208.07 | 1 | Carbamazepine-$^{13}C_6$ |
| Oxybenzone | 131-57-7 | [M+H] | 229.0864 | 10.9 | 151.04 | 1 | Oxybenzone-$d_3$ |
| Paraxanthine | 611-59-6 | [M+H] | 181.0726 | 2.8 | 124.05 | 5 | Caffeine-$^{13}C_3$ |
| Penciclovir | 39809-25-1 | [M+H] | 254.1248 | 1.3 | 152.06 | 25 | Allopurinol-$d_2$ |
| Pentoxyfylline | 6493-05-6 | [M+H] | 279.1457 | 4.7 | 181.07 | 1 | Carbofuran-$d_3$ |
| Perfluorobutanoic Acid | 375-22-4 | [M-H] | 212.9792 | 4.5 | 168.99 | 100 | Sucralose-$d_6$ |
| Perfluorooctanoic Acid | 335-67-1 | [M-H] | 412.9665 | 12.0 | 218.99 | 25 | Sucralose-$d_6$ |
| Phenobarbital | 50-06-6 | [M-H] | 231.0770 | 5.3 | 188.07 | 25 | Sucralose-$d_6$ |
| Phenytoin | 57-41-0 | [M+H] | 253.0977 | 6.5 | 182.10 | 5 | Carbofuran-$d_3$ |
| Pirimicarb | 23103-98-2 | [M+H] | 239.1508 | 4.1 | 182.13 | 1 | Pirimicarb-$d_6$ |
| Pirimiphos-ethyl | 23505-41-1 | [M+H] | 334.1349 | 12.9 | 198.11 | 1 | Ibuprofen-$d_3$ |
| Primidone | 125-33-7 | [M+H] | 219.1128 | 4.6 | 162.09 | 5 | Pirimicarb-$d_6$ |
| Progesterone | 57-83-0 | [M+H] | 315.2324 | 11.6 | 297.22 | 5 | Azoxystrobin-$d_4$ |
| Prohexadione | 127277-53-6 | [M-H] | 211.0606 | 5.5 | 125.02 | 5 | Sucralose-$d_6$ |
| Prometon | 1610-18-0 | [M+H] | 226.1662 | 5.8 | 184.12 | 1 | Carbaryl-$d_7$ |
| Propachlor | 1918-16-7 | [M+H] | 212.0836 | 7.5 | 170.04 | 1 | Isoproturon-$d_6$ |
| Propachlor-ESA | 123732-85-4 | [M-H] | 256.0638 | 4.7 | 162.02 | 5 | Sucralose-$d_6$ |
| Propachlor-OXA | 70628-36-3 | [M+H] | 208.0968 | 4.8 | 120.04 | 5 | Carbofuran-$d_3$ |
| Propazine | 139-40-2 | [M+H] | 230.1167 | 8.6 | 146.02 | 1 | Azoxystrobin-$d_4$ |
| Propoxur | 114-26-1 | [M+H] | 210.1125 | 6.0 | 111.04 | 25 | Carbofuran-$d_3$ |
| Propranolol | 525-66-6 | [M+H] | 260.1651 | 5.9 | 183.08 | 1 | Dextromethorphan-$d_3$ |
| Propyzamide | 23950-58-5 | [M+H] | 256.0296 | 9.6 | 172.96 | 1 | Isoproturon-$d_6$ |
| Pseudoephedrine | 299-42-3 | [M+H] | 166.1229 | 2.8 | 133.09 | 1 | Iopromide-$d_3$ |
| Pyrazophos | 13457-18-6 | [M+H] | 374.0934 | 12.1 | 222.09 | 1 | Ibuprofen-$d_3$ |
| Ranitidine | 66357-35-5 | [M+H] | 315.1485 | 2.2 | 176.05 | 5 | Ranitidine-$d_6$ |
| Ritalinic Acid | 19395-41-6 | [M+H] | 220.1332 | 4.1 | 174.13 | 1 | Thiabendazole-$d_4$ |
| Serotonin | 153-98-0 | [M+H] | 177.1022 | 1.8 | 146.06 | 5 | Morphine-$d_3$ |
| Sertraline | 79617-96-2 | [M+H] | 306.0816 | 8.9 | 275.04 | 25 | Sertraline-$d_3$ |
| Siduron | 1982-49-6 | [M+H] | 233.1654 | 8.9 | 137.07 | 5 | Naproxen-methoxy-$d_3$ |
| Simazine | 122-34-9 | [M+H] | 202.0854 | 6.0 | 132.03 | 1 | Atrazine-$d_5$ |

**Table A5** (Continued)

| Micropollutant | CAS number | Adduct | Exact Mass | RT | Diagnostic Fragment | LOQ | Internal Standard Assignment |
|---|---|---|---|---|---|---|---|
| Sitagliptin | 486460-32-6 | [M+H] | 408.1259 | 4.6 | 235.08 | 5 | Dimethoate-$d_6$ |
| Sucralose | 56038-13-2 | [M+FA-H] | 447.0499 | 3.9 | 278.15 | 5 | Sucralose-$d_6$ |
| Sulfadimethoxine | 122-11-2 | [M+H] | 311.0808 | 4.9 | 237.04 | 1 | Sulfadimethoxine-$d_6$ |
| Sulfamethazine | 57-68-1 | [M+H] | 279.0910 | 3.3 | 204.04 | 5 | Sulfadimethoxine-$d_6$ |
| Sulfamethoxazole | 723-46-6 | [M+H] | 254.0594 | 3.9 | 156.01 | 1 | Sulfamethoxazole-phenyl-$^{13}C_6$ |
| Sulfathiazole | 72-14-0 | [M+H] | 256.0209 | 2.6 | 156.01 | 25 | Sulfadimethoxine-$d_6$ |
| TDCPP | 13674-87-8 | [M+H] | 428.8917 | 11.9 | 98.98 | 25 | Tris-2-chloroethyl-phosphate-$d_{12}$ |
| Temazepam | 846-50-4 | [M+H] | 301.0744 | 8.1 | 255.07 | 1 | Isoproturon-$d_6$ |
| Terbutylazine | 5915-41-3 | [M+H] | 230.1167 | 8.9 | 174.05 | 1 | Naproxen-methoxy-$d_3$ |
| Testosterone | 58-22-0 | [M+H] | 289.2168 | 9.6 | 253.20 | 1 | Azoxystrobin-$d_4$ |
| Thiabendazole | 148-79-8 | [M+H] | 202.0433 | 3.5 | 175.03 | 1 | Thiabendazole-$d_4$ |
| Triamterene | 396-01-0 | [M+H] | 254.1154 | 4.0 | 237.09 | 1 | Thiabendazole-$d_4$ |
| Tributyl-phosphate | 126-73-8 | [M+H] | 267.1725 | 13.1 | 155.05 | 25 | Ibuprofen-$d_3$ |
| Triclosan | 3380-34-5 | [M-H] | 286.9428 | 13.5 | 165.89 | 50 | Triclosan-$d_3$ |
| Trimethoprim | 738-70-5 | [M+H] | 291.1451 | 3.4 | 245.10 | 1 | Trimethoprim-$d_9$ |
| Trinexapac-ethyl | 95266-40-3 | [M+H] | 253.1070 | 7.9 | 183.03 | 1 | Isoproturon-$d_6$ |
| Tris(2-chloroethyl)phosphate | 115-96-8 | [M+H] | 284.9617 | 6.3 | 160.98 | 5 | Tris-2-chloroethyl-phosphate-$d_{12}$ |
| Valsartan | 137862-53-4 | [M+H] | 436.2348 | 9.7 | 235.10 | 1 | Isoproturon-$d_6$ |
| Venlafaxine | 93413-69-5 | [M+H] | 278.2114 | 5.5 | 121.06 | 1 | Venlafaxine-$d_6$ |
| Verapamil | 52-53-9 | [M+H] | 455.2910 | 6.4 | 303.21 | 1 | Diltiazem-$d_4$ |
| Warfarin | 2610-86-8 | [M+H] | 309.1121 | 9.5 | 251.07 | 1 | Azoxystrobin-$d_4$ |

**Table A6:** Isotope-labeled internal reference standard information.

| Micropollutant | CAS number | Supplier | Adduct | Exact Mass (m/z) | RT (min) |
|---|---|---|---|---|---|
| 2-methyl-3-isothiazolinone-$d_3$ | 1329509-49-0 | TRC | [M+H] | 119.0353 | 1.5 |
| Acetaminophen-$d_4$ | 64315-36-2 | Sigma Aldrich | [M+H] | 156.0963 | 2.3 |
| Allopurinol-$d_2$ | 916979-34-5 | TRC | [M+H] | 139.0583 | 1.5 |
| Atenolol-$d_7$ | 1202864-50-3 | Sigma Aldrich | [M+H] | 274.2148 | 2.1 |
| Atrazine-$d_5$ | 163165-75-1 | Sigma Aldrich | [M+H] | 221.1330 | 7.2 |
| Azoxystrobin-$d_4$ | 1346606-39-0 | TRC | [M+H] | 408.1492 | 8.7 |
| Caffeine-$^{13}C_3$ | n/a | Sigma Aldrich | [M+H] | 198.0977 | 3.4 |
| Carbamazepine-$^{13}C_6$ | n/a | Sigma Aldrich | [M+H] | 243.1224 | 6.7 |
| Carbaryl-$d_7$ | 362049-56-7 | TRC | [M+H] | 209.1302 | 6.5 |
| Carbofuran-$d_3$ | 1007459-98-4 | Sigma Aldrich | [M+H] | 225.1313 | 6.1 |
| Celecoxib-$d_4$ | 544686-20-6 | TRC | [M+H] | 386.1083 | 10.9 |
| Cimetidine-$d_3$ | 1185237-29-9 | TRC | [M+H] | 256.1418 | 2.3 |
| Citalopram-$d_6$ | 1246819-94-2 | TRC | [M+H] | 331.2087 | 5.7 |
| Dextromethorphan-$d_3$ | 524713-56-2 | TRC | [M+H] | 275.2197 | 5.7 |
| Diclofenac-$13C_6$ | 1261393-73-0 | Sigma Aldrich | [M+H] | 302.0441 | 11.5 |
| Diltiazem-$d_4$ | 1217769-52-2 | TRC | [M+H] | 419.1937 | 6.2 |
| Dimethoate-$d_6$ | 1219794-81-6 | TRC | [M+H] | 236.0446 | 4.5 |
| Erythromycin-$^{13}C$-$d_3$ | 959119-26-7 | TRC | [M+H] | 738.4907 | 7.1 |
| Fexofenadine-$d_6$ | 548783-71-7 | TRC | [M+H] | 508.3328 | 7.1 |
| Fluoxetine-$d_5$ | 1173020-43-3 | Sigma Aldrich | [M+H] | 315.1727 | 8.1 |
| Gembifrozil-$d_6$ | n/a | TRC | [M+H] | 257.2024 | 13.4 |
| Ibuprofen-$d_3$ | 121662-14-4 | Sigma Aldrich | [M+Na] | 232.1387 | 11.9 |
| Imidacloprid-$d_4$ | 1015855-75-0 | Sigma Aldrich | [M+H] | 260.0852 | 4.0 |
| Iodocarb-d9 | 1246815-08-6 | TRC | [M+H] | 291.0550 | 7.3 |
| Iopromide-$d_3$ | 1189947-73-6 | TRC | [M+H] | 794.8959 | 2.5 |
| Isoproturon-$d_6$ | 217487-17-7 | Sigma Aldrich | [M+H] | 213.1869 | 7.6 |
| Mecoprop-$d_3$ | 352431-15-3 | TRC | [M-H] | 216.0501 | 9.7 |
| Metformin-$d_6$ | 1185166-01-1 | TRC | [M+H] | 136.1464 | 1.1 |
| Metoprolol Acid-$d_5$ | 1215404-47-9 | TRC | [M+H] | 273.1863 | 3.2 |
| Morphine-$d_3$ | 118357-24-7 | Sigma Aldrich | [M+H] | 289.1626 | 1.6 |
| Naproxen-methoxy-$d_3$ | 958293-79-3 | Sigma Aldrich | [M+H] | 234.1204 | 8.9 |
| Oxybenzone-$d_3$ | n/a | TRC | [M+H] | 232.1053 | 10.8 |
| Pirimicarb-$d_6$ | 1015854-66-6 | Sigma Aldrich | [M+H] | 245.1885 | 4.1 |
| rac-Efavirenz-$d_4$ | 1246812-58-7 | TRC | [M+H] | 320.0598 | 11.5 |
| Ranitidine-$d_6$ | 1185238-09-8 | TRC | [M+H] | 321.1862 | 2.1 |
| Sertraline-$d_3$ | 79559-97-0 | TRC | [M+H] | 309.0999 | 8.6 |
| Sucralose-$d_6$ | n/a | TRC | [M+FA-H] | 447.0499 | 3.9 |
| Sulfadimethoxine-$d_6$ | 73068-02-7 | Sigma Aldrich | [M+H] | 317.1185 | 4.9 |
| Sulfamethoxazole-phenyl-$^{13}C_6$ | 1196157-90-0 | Sigma Aldrich | [M+H] | 260.0795 | 3.9 |
| Thiabendazole-$d_4$ | 1190007-20-5 | TRC | [M+H] | 206.0685 | 3.4 |
| Triclosan-$d_3$ | 1020719-98-5 | TRC | [M-H] | 289.9616 | 13.5 |
| Trimethoprim-$d_9$ | 1189460-62-5 | Sigma Aldrich | [M+H] | 300.2016 | 3.2 |
| Tris-2-chloroethyl-phosphate-$d_{12}$ | 1276500-47-0 | TRC | [M+H] | 297.0365 | 6.3 |
| Venlafaxine-$d_6$ | 1062606-12-5 | Sigma Aldrich | [M+H] | 284.2491 | 5.3 |

TRC = Toronto research chemicals.

## A.4 Geospatial analysis

**Table A7:** List of geospatial data sources.

| Spatial Reference | Purpose | Date | Source |
|---|---|---|---|
| Area hydrography | Hydraulic distances | 2010 | U.S. Census[4] |
| Population data | Population information | 2010 | U.S. Census[5] |
| State borders | Background visuals | 2010 | U.S. Census[6] |
| Digital elevation models | Watershed delineations | 2015 | U.S. Geological Survey[7] |
| Land cover | Watershed features | 2011 | U.S. Geological Survey[8] |
| EPA facility registry service | Watershed features | 2014 | U.S. Environmental Protection Agency[9] |

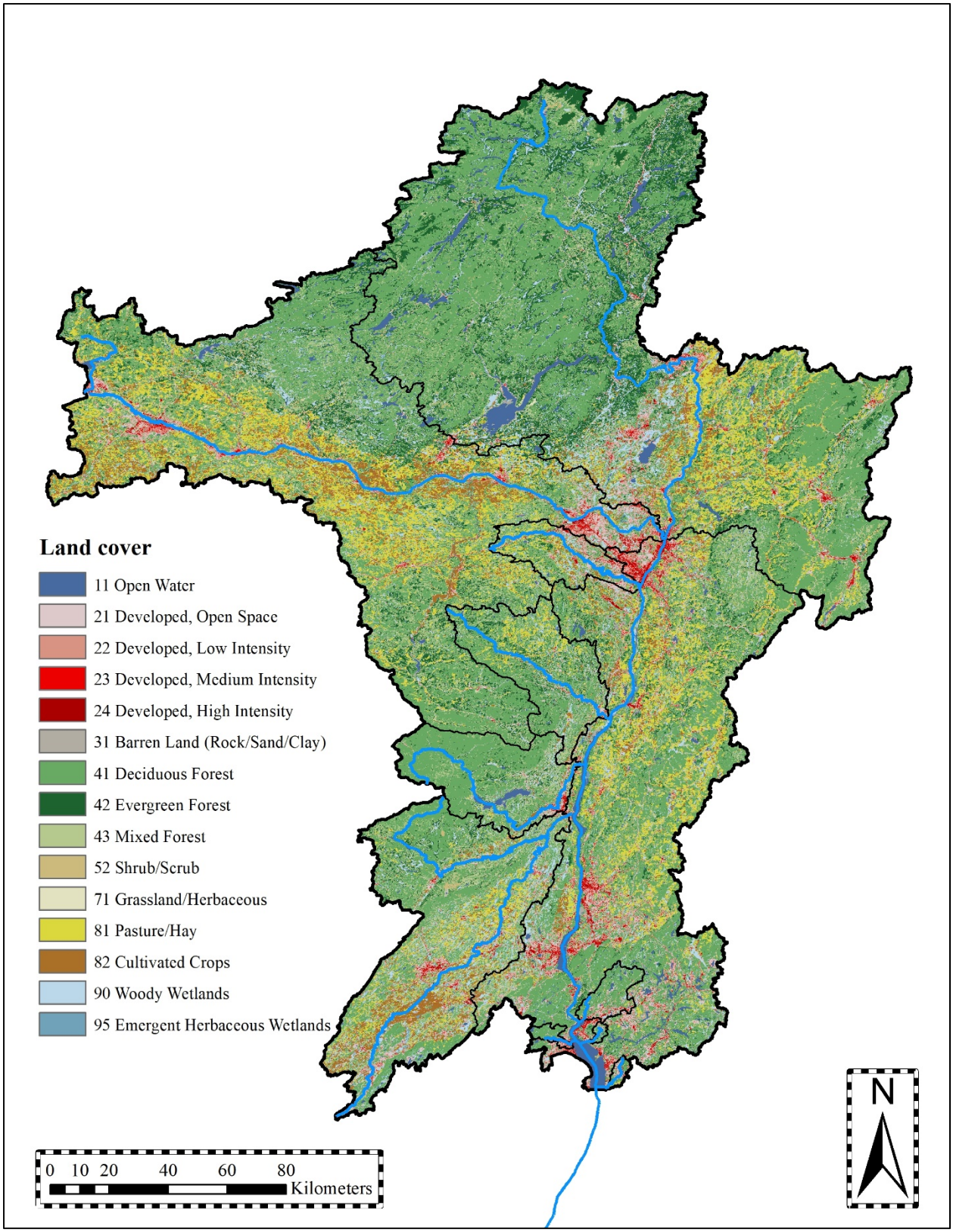



**Figure A1:** Land cover in the Hudson River Estuary catchment area.

**Table A8:** Percent land cover in the Hudson River Estuary (HRE) and each sub-watershed.

| Watershed | Area ($km^2$) | Open Water | Developed Open Space | Developed Low Intensity | Developed Medium Intensity | Developed High Intensity | Barren Land | Deciduous Forest | Evergreen Forest | Mixed Forest | Shrub/ Scrub | Herb. | Hay/ Pasture | Cultivated Crops | Woody Wetlands | Em. Herb. Wetlands |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| HRE | 34300 | 2.9% | 5.4% | 2.3% | 1.1% | 0.4% | 0.2% | 41.2% | 11.6% | 9.9% | 2.1% | 0.6% | 10.5% | 4.9% | 6.0% | 0.7% |
| PR | 37 | 0.8% | 28.9% | 8.5% | 5.4% | 1.2% | 0.0% | 41.8% | 6.7% | 0.5% | 0.3% | 0.0% | 4.7% | 0.1% | 1.1% | 0.1% |
| CB | 49 | 3.2% | 10.7% | 6.3% | 4.1% | 0.7% | 0.0% | 68.7% | 0.2% | 0.2% | 0.2% | 0.7% | 0.2% | 0.3% | 2.6% | 2.0% |
| FB | 20 | 1.2% | 17.1% | 6.7% | 2.9% | 0.6% | 0.0% | 57.1% | 5.0% | 0.6% | 0.2% | 0.0% | 1.6% | 0.0% | 6.7% | 0.4% |
| AC | 183 | 3.1% | 15.5% | 5.1% | 2.7% | 0.8% | 0.0% | 62.6% | 2.8% | 0.7% | 0.5% | 0.0% | 1.5% | 0.1% | 4.0% | 0.6% |
| RC | 3224 | 1.3% | 7.2% | 2.3% | 0.9% | 0.3% | 0.1% | 38.4% | 3.7% | 13.4% | 0.6% | 0.3% | 9.7% | 10.4% | 10.4% | 0.9% |
| EC | 1097 | 2.9% | 4.0% | 1.4% | 0.5% | 0.3% | 0.4% | 61.7% | 6.4% | 14.8% | 0.2% | 0.1% | 0.8% | 1.9% | 4.4% | 0.1% |
| CC | 1067 | 0.8% | 5.6% | 1.2% | 0.3% | 0.1% | 0.0% | 47.0% | 12.4% | 14.0% | 0.6% | 0.5% | 9.4% | 2.2% | 5.7% | 0.2% |
| NK | 465 | 0.7% | 12.0% | 6.2% | 2.4% | 1.1% | 0.0% | 31.1% | 6.0% | 10.9% | 0.7% | 0.3% | 13.5% | 6.4% | 8.2% | 0.5% |
| MR | 8965 | 1.7% | 4.4% | 2.1% | 0.9% | 0.3% | 0.1% | 35.2% | 10.0% | 10.2% | 3.8% | 1.5% | 15.6% | 8.7% | 5.0% | 0.5% |
| UHR | 11999 | 3.5% | 3.1% | 1.3% | 0.6% | 0.2% | 0.2% | 43.6% | 19.5% | 11.1% | 1.6% | 0.2% | 6.3% | 2.4% | 5.9% | 0.6% |

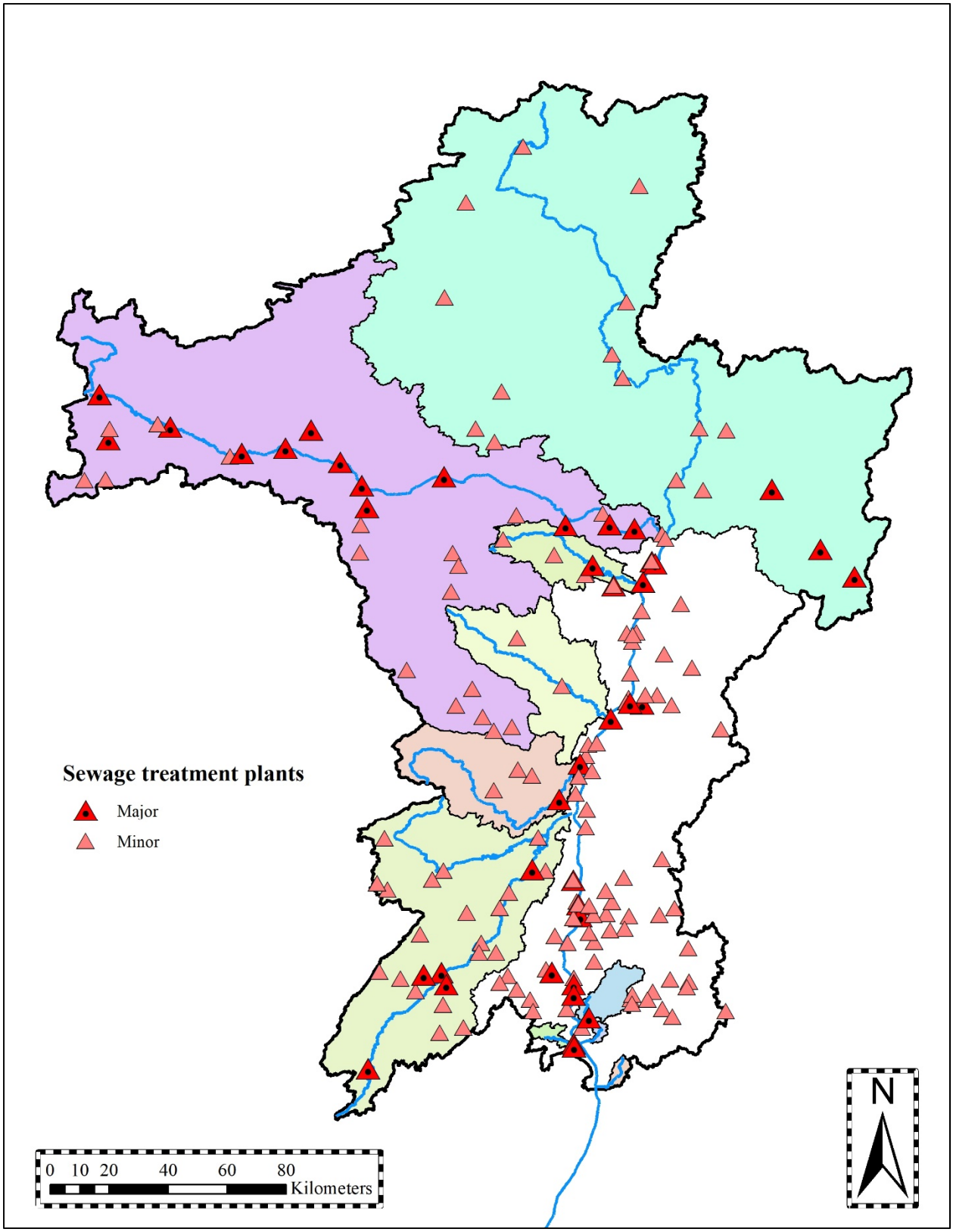



**Figure A2:** Sewage treatment plants in the Hudson River Estuary catchment area.

## A.5 Micropollutant detection frequencies and concentrations

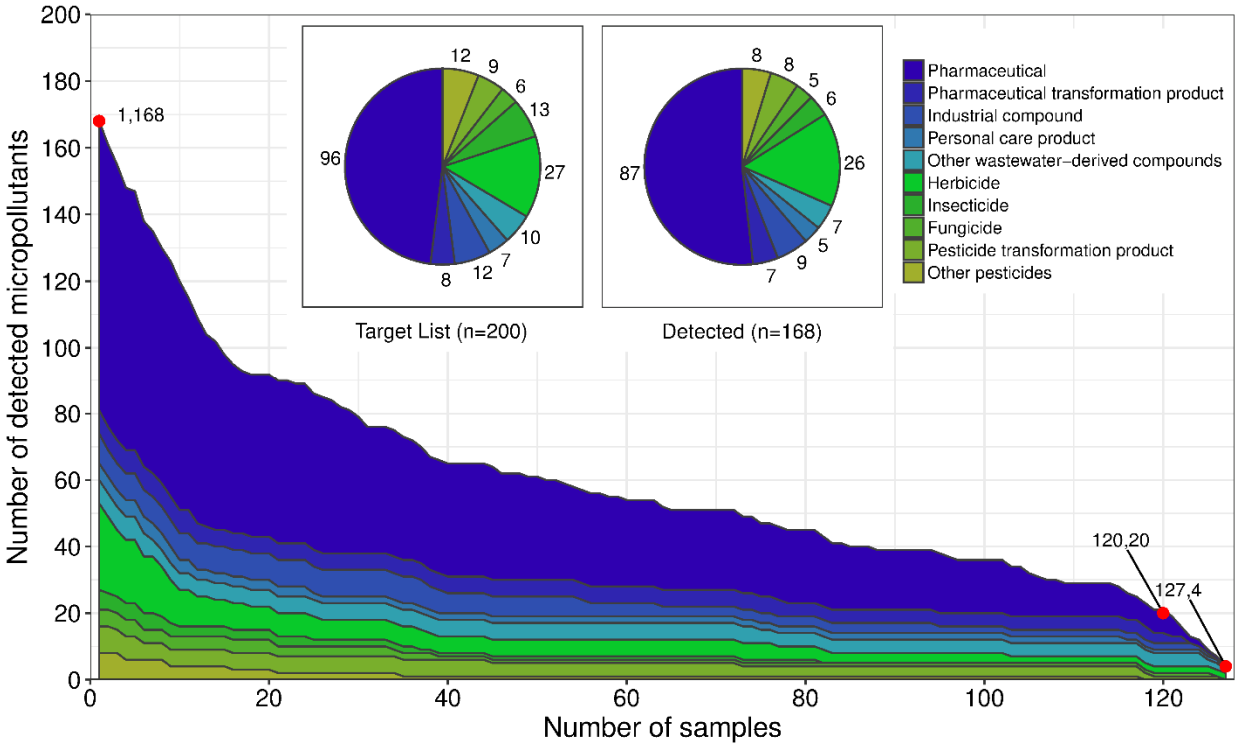



**Figure A3:** The number of detected micropollutants in at least n samples by use-class. Blue shaded use-classes are considered to be wastewater-derived; green shaded are considered to be agricultural-derived.

**Table A9:** Micropollutant frequency of detection, minimum, median, and maximum concentrations, and spatiotemporal cluster group.

| Compound | Frequency of Detection (n = 127) | Minimum concentration[a] ($ng \cdot L^{-1}$) | Median concentration[b] ($ng \cdot L^{-1}$) | Maximum concentration[b] ($ng \cdot L^{-1}$) | Occurrence cluster | Occurrence sub-cluster | Normalized concentration sub-cluster |
|---|---|---|---|---|---|---|---|
| 10,11-dihydrocarbamazepine | 11 | 1 | 4 | 10 | Diffuse | Diffuse-specific | |
| 2,4-D | 102 | 5 | 12 | 108 | Core | | D |
| 2,6-dichlorobenzamide | 0 | NA | NA | NA | | | |
| 2,6-dimethoxyphenol | 0 | NA | NA | NA | | | |
| 2-aminobenzimidazole | 5 | 5 | 6 | 9 | Diffuse | Diffuse-specific | |
| 2-ethyl-2-phenyl-malonamide | 12 | 6 | 38 | 184 | STP | STP-exclusive | |
| 2-methylisothiazolin-3-one (MI) | 125 | 2 | 16 | 67 | Core | | C |
| 6-benzylaminopurine | 15 | 2 | 16 | 94 | STP | STP-exclusive | |
| Abacavir | 11 | 7 | 12 | 203 | STP | STP-exclusive | |
| Abscisic Acid | 0 | NA | NA | NA | | | |
| Acebutolol | 27 | 1 | 4 | 25 | STP | WW-sources | |
| Acephate | 9 | 7 | 23 | 39 | Diffuse | Diffuse-general | |
| Acesulfame | 124 | 11 | 62 | *12945* | Core | | B |
| Acetaminophen | 14 | 19 | 41 | 342 | Diffuse | Diffuse-general | |
| Acetamiprid | 3 | 5 | 6 | 14 | Diffuse | Diffuse-specific | |
| Acetazolamide | 16 | 7 | 22 | 145 | STP | STP-exclusive | |
| Acetochlor/Alachlor | 25 | 86 | 207 | 383 | Diffuse | Diffuse-general | |
| Adrenalone | 10 | 42 | 260 | *2247* | Diffuse | Diffuse-general | |
| Adrenosterone | 0 | NA | NA | NA | | | |
| Albuterol | 45 | 2 | 5 | 43 | Core | | D |
| Aldicarb | 0 | NA | NA | NA | | | |
| Amcinonide | 3 | 5 | 6 | 6 | Diffuse | Diffuse-general | |
| Amisulpride | 10 | 1 | 3 | 6 | Diffuse | Diffuse-specific | |
| Amitriptyline | 36 | 1 | 3 | 36 | STP | WW-sources | |
| Amphetamine | 3 | 6 | 21 | 22 | Diffuse | Diffuse-general | |
| Ampicillin | 0 | NA | NA | NA | | | |
| Arecoline | 4 | 8 | 15 | 82 | STP | STP-exclusive | |
| Atenolol | 116 | 1 | 8 | 992 | Core | | B |
| Atenolol Acid | 122 | 8 | 66 | *23774* | Core | | B |
| Atomoxetine | 13 | 1 | 2 | 15 | Diffuse | Diffuse-general | |
| Atorvastatin | 10 | 136 | *1213* | *6888* | STP | STP-exclusive | |
| Atrazine | 127 | 1 | 14 | 204 | Core | | A |
| Atrazine-2-hydroxy | 126 | 2 | 9 | 88 | Core | | A |
| Atrazine-desethyl | 117 | 2 | 11 | 81 | Core | | A |
| Atrazine-desethyl-desisopropyl | 44 | 25 | 54 | 326 | Core | | A |

**Table A9** (Continued)

| Compound | Frequency of Detection (n = 127) | Minimum concentration[a] ($ng \cdot L^{-1}$) | Median concentration[b] ($ng \cdot L^{-1}$) | Maximum concentration[b] ($ng \cdot L^{-1}$) | Occurrence cluster | Occurrence sub-cluster | Normalized concentration sub-cluster |
|---|---|---|---|---|---|---|---|
| Atropine | 27 | 1 | 5 | 16 | Diffuse | Diffuse-general | |
| Azoxystrobin | 38 | 1 | 2 | 10 | Diffuse | Diffuse-general | |
| Baclofen | 14 | 5 | 19 | 158 | STP | STP-exclusive | |
| Bendiocarb | 0 | NA | NA | NA | | | |
| Bentazon | 9 | 5 | 25 | 336 | Diffuse | Diffuse-general | |
| Benzisothiazolin-3-one (BIT) | 8 | 6 | 9 | 16 | Diffuse | Diffuse-general | |
| Benzotriazole | 123 | 2 | 88 | *2521* | Core | | B |
| Benzotriazole-methyl-1H | 125 | 1 | 131 | *1931* | Core | | B |
| Benzoylecgonine | 74 | 1 | 7 | 437 | Core | | D |
| Bromacil | 0 | NA | NA | NA | | | |
| Bupropion | 72 | 1 | 2 | 660 | Core | | B |
| Butalbital | 0 | NA | NA | NA | | | |
| Butocarboxim | 0 | NA | NA | NA | | | |
| Caffeine | 104 | 26 | 47 | *2339* | Core | | C |
| Candesartan | 5 | 2 | 3 | 4 | Diffuse | Diffuse-specific | |
| Carbamazepine | 108 | 1 | 6 | 262 | Core | | B |
| Carbaryl | 81 | 1 | 6 | 33 | Core | | D |
| Carbazole | 0 | NA | NA | NA | | | |
| Carbendazim | 117 | 2 | 6 | 207 | Core | | B |
| Carbofuran | 0 | NA | NA | NA | | | |
| Carisoprodol | 5 | 6 | 18 | 30 | Diffuse | Diffuse-general | |
| Celecoxib | 22 | 5 | 12 | 72 | STP | WW-sources | |
| Chloridazon | 7 | 2 | 6 | 12 | Diffuse | Diffuse-specific | |
| Chlorpheniramine | 6 | 2 | 5 | 11 | Diffuse | Diffuse-general | |
| Cimetidine | 5 | 7 | 24 | 244 | STP | STP-exclusive | |
| Ciprofloxacin | 12 | 8 | 46 | 207 | STP | WW-sources | |
| cis-Diltiazem | 34 | 1 | 6 | 179 | STP | WW-sources | |
| Citalopram | 35 | 1 | 4 | 56 | STP | WW-sources | |
| Clarithromycin | 28 | 6 | 31 | 266 | STP | WW-sources | |
| Climbazole | 11 | 1 | 3 | 7 | Diffuse | Diffuse-general | |
| Clofibric Acid | 2 | 5 | 7 | 8 | Diffuse | Diffuse-specific | |
| Codeine | 26 | 1 | 8 | 112 | STP | WW-sources | |
| Corticosterone | 0 | NA | NA | NA | | | |
| Cotinine | 126 | 1 | 9 | 444 | Core | | C |

**Table A9** (Continued)

| Compound | Frequency of Detection (n = 127) | Minimum concentration[a] ($ng \cdot L^{-1}$) | Median concentration[b] ($ng \cdot L^{-1}$) | Maximum concentration[b] ($ng \cdot L^{-1}$) | Occurrence cluster | Occurrence sub-cluster | Normalized concentration sub-cluster |
|---|---|---|---|---|---|---|---|
| Coumarin | 2 | 26 | 34 | 42 | Diffuse | Diffuse-general | |
| Cyanazine | 2 | 117 | 129 | 141 | | | |
| Cyflufenamid | 0 | NA | NA | NA | | | |
| DEET | 126 | 2 | 26 | *3910* | Core | | C |
| Dehydroacetic Acid | 5 | 51 | 76 | 184 | Diffuse | Diffuse-general | |
| Desvenlafaxine | 122 | 2 | 13 | *1636* | Core | | B |
| Dexamethasone | 0 | NA | NA | NA | | | |
| Dextromethorphan | 74 | 1 | 2 | 90 | Core | | B |
| Diclofenac | 11 | 26 | 90 | 856 | STP | STP-exclusive | |
| Diethyl phthalate | 63 | 25 | 47 | 347 | Core | | A |
| Dikegulac | 0 | NA | NA | NA | | | |
| Dimethachlor | 1 | 1 | 1 | 1 | Diffuse | Diffuse-general | |
| Dimethoate | 1 | 3 | 3 | 3 | Diffuse | Diffuse-specific | |
| Diphenhydramine | 104 | 1 | 4 | *1085* | Core | | B |
| Diuron | 20 | 5 | 13 | 90 | STP | STP-exclusive | |
| Efavirenz | 0 | NA | NA | NA | | | |
| Estriol | 0 | NA | NA | NA | | | |
| Estrone | 0 | NA | NA | NA | | | |
| Ethofumesate | 1 | 48 | 48 | 48 | Diffuse | Diffuse-general | |
| Ethyl butylacetylaminopropionate | 24 | 5 | 20 | 626 | STP | WW-sources | |
| Famciclovir | 7 | 6 | 8 | 15 | Diffuse | Diffuse-specific | |
| Famotidine | 3 | 99 | 153 | 396 | STP | STP-exclusive | |
| Fexofenadine | 121 | 5 | 40 | *6333* | Core | | B |
| Fluconazole | 82 | 1 | 3 | 156 | Core | | B |
| Flucytosine | 0 | NA | NA | NA | | | |
| Fluoxetine | 16 | 5 | 11 | 47 | STP | WW-sources | |
| Folic Acid | 0 | NA | NA | NA | | | |
| Furosemid | 14 | 28 | 148 | 906 | STP | STP-exclusive | |
| Gabapentin | 127 | 3 | 57 | *6784* | Core | | B |
| Gemfibrozil | 30 | 11 | 93 | *1949* | STP | WW-sources | |
| Gibberellic Acid | 34 | 27 | 127 | *14287* | Diffuse | Diffuse-general | |
| Hexamethylphosphoramide | 1 | 1 | 1 | 1 | Diffuse | Diffuse-general | |
| Hexazinone | 2 | 8 | 12 | 16 | Diffuse | Diffuse-general | |
| Hydrochlorothiazide | 9 | 88 | 307 | *1340* | STP | STP-exclusive | |

**Table A9** (Continued)

| Compound | Frequency of Detection (n = 127) | Minimum concentration[a] ($ng \cdot L^{-1}$) | Median concentration[b] ($ng \cdot L^{-1}$) | Maximum concentration[b] ($ng \cdot L^{-1}$) | Occurrence cluster | Occurrence sub-cluster | Normalized concentration sub-cluster |
|---|---|---|---|---|---|---|---|
| Hydrocodone | 12 | 7 | 13 | 116 | STP | STP-exclusive | |
| Hydrocortisone | 0 | NA | NA | NA | | | |
| Ibuprofen | 12 | 33 | 379 | *3273* | STP | WW-sources | |
| Imidacloprid | 36 | 5 | 12 | 103 | STP | WW-sources | |
| Iopromide | 12 | 59 | 267 | *17337* | STP | STP-exclusive | |
| Ioxynil | 9 | 1 | 3 | 27 | Diffuse | Diffuse-specific | |
| Irbesartan | 84 | 1 | 3 | 87 | Core | | C |
| Isophorone Diisocyanate | 0 | NA | NA | NA | | | |
| Isoproturon | 13 | 1 | 2 | 4 | Diffuse | Diffuse-general | |
| Ketamine | 29 | 1 | 3 | 128 | STP | WW-sources | |
| Ketoprofen | 57 | 1 | 2 | 12 | Core | | D |
| Lamotrigine | 124 | 1 | 29 | *1699* | Core | | C |
| Levetiracetam | 105 | 3 | 15 | *5466* | Core | | C |
| Lidocaine | 115 | 2 | 9 | 1151 | Core | | B |
| Linuron | 11 | 5 | 8 | 13 | Diffuse | Diffuse-general | |
| Losartan | 87 | 2 | 11 | 519 | Core | | C |
| Mabuterol | 1 | 15 | 15 | 15 | Diffuse | Diffuse-general | |
| Malaoxon | 0 | NA | NA | NA | | | |
| MCPA | 8 | 5 | 7 | 9 | Diffuse | Diffuse-specific | |
| Mecoprop | 37 | 5 | 9 | 29 | STP | WW-sources | |
| Melamine | 72 | 251 | *1022* | *7585* | Core | | D |
| Meprobamate | 10 | 8 | 30 | 89 | STP | STP-exclusive | |
| Metalaxyl | 33 | 2 | 7 | 41 | Core | | A |
| Metamitron | 8 | 6 | 9 | 12 | Diffuse | Diffuse-specific | |
| Metaxalone | 45 | 1 | 4 | 30 | STP | WW-sources | |
| Metformin | 120 | 44 | 296 | *112574* | Core | | B |
| Methadone | 63 | 1 | 2 | 103 | STP | WW-sources | |
| Methocarbamol | 96 | 5 | 11 | *1646* | Core | | B |
| Methomyl | 0 | NA | NA | NA | | | |
| Metolachlor | 127 | 2 | 15 | 298 | Core | | A |
| Metolachlor-ESA | 118 | 11 | 36 | 279 | Core | | D |
| Metoprolol | 121 | 2 | 12 | *3413* | Core | | B |
| Metribuzin | 7 | 1 | 1 | 3 | Diffuse | Diffuse-general | |
| Metsulfuron-methyl | 3 | 36 | 114 | 154 | Diffuse | Diffuse-general | |

**Table A9** (Continued)

| Compound | Frequency of Detection (n = 127) | Minimum concentration[a] ($ng \cdot L^{-1}$) | Median concentration[b] ($ng \cdot L^{-1}$) | Maximum concentration[b] ($ng \cdot L^{-1}$) | Occurrence cluster | Occurrence sub-cluster | Normalized concentration sub-cluster |
|---|---|---|---|---|---|---|---|
| Molinate | 0 | NA | NA | NA | | | |
| Morphine | 15 | 5 | 51 | 405 | STP | STP-exclusive | |
| N,N-didesmethyl venlafaxine | 0 | NA | NA | NA | | | |
| N4-acetylsulfamethoxazole | 77 | 1 | 5 | 446 | Core | | B |
| Nadolol | 30 | 2 | 7 | 129 | STP | WW-sources | |
| Naproxen | 59 | 10 | 28 | *2243* | Core | | D |
| Nicotine | 124 | 5 | 27 | *2989* | Core | | C |
| Norfloxacin | 9 | 7 | 13 | 24 | STP | WW-sources | |
| Ofloxacin | 24 | 6 | 24 | 195 | STP | WW-sources | |
| Oxazepam | 14 | 5 | 21 | 232 | STP | STP-exclusive | |
| Oxcarbazepine | 29 | 1 | 9 | 256 | STP | WW-sources | |
| Oxybenzone | 115 | 1 | 12 | 404 | Core | | C |
| Paraxanthine/Theophylline | 122 | 5 | 16 | 974 | Core | | C |
| Penciclovir | 53 | 25 | 64 | *1226* | Diffuse | Diffuse-general | |
| Pentoxyfylline | 10 | 1 | 5 | 19 | Diffuse | Diffuse-specific | |
| Perfluorobutanoic Acid | 39 | 151 | *1850* | *9996* | Diffuse | Diffuse-general | |
| Perfluorooctanoic Acid | 94 | 2 | 26 | 338 | Core | | D |
| Phenobarbital | 0 | NA | NA | NA | | | |
| Phenytoin | 1 | 12 | 12 | 12 | Diffuse | Diffuse-general | |
| Pirimicarb | 0 | NA | NA | NA | | | |
| Pirimiphos-ethyl | 24 | 1 | 2 | 3 | Diffuse | Diffuse-general | |
| Primidone | 5 | 5 | 22 | 118 | Diffuse | Diffuse-general | |
| Progesterone | 6 | 5 | 6 | 7 | Diffuse | Diffuse-specific | |
| Prohexadione | 0 | NA | NA | NA | | | |
| Prometon | 76 | 1 | 3 | 8 | Core | | D |
| Propachlor | 5 | 1 | 4 | 8 | Diffuse | Diffuse-general | |
| Propachlor-ESA | 2 | 8 | 9 | 10 | Diffuse | Diffuse-specific | |
| Propachlor-OXA | 5 | 19 | 44 | 260 | Diffuse | Diffuse-general | |
| Propazine | 17 | 1 | 1 | 3 | Core | | |
| Propoxur | 7 | 29 | 148 | 311 | STP | STP-exclusive | |
| Propranolol | 48 | 1 | 4 | 125 | STP | WW-sources | |
| Propyzamide | 1 | 2 | 2 | 2 | Diffuse | Diffuse-general | |
| Pseudoephedrine | 95 | 1 | 3 | 376 | Core | | C |
| Pyrazophos | 20 | 1 | 5 | 17 | Diffuse | Diffuse-general | |

**Table A9** (Continued)

| Compound | Frequency of Detection (n = 127) | Minimum concentration[a] ($ng \cdot L^{-1}$) | Median concentration[b] ($ng \cdot L^{-1}$) | Maximum concentration[b] ($ng \cdot L^{-1}$) | Occurrence cluster | Occurrence sub-cluster | Normalized concentration sub-cluster |
|---|---|---|---|---|---|---|---|
| Ranitidine | 7 | 39 | 511 | 951 | STP | STP-exclusive | |
| Ritalinic Acid | 118 | 1 | 5 | 441 | Core | | B |
| Serotonin | 102 | 5 | 9 | 434 | Core | | C |
| Sertraline | 3 | 26 | 39 | 88 | Diffuse | Diffuse-general | |
| Siduron | 5 | 6 | 8 | 12 | Diffuse | Diffuse-specific | |
| Simazine | 82 | 1 | 5 | 53 | Core | | D |
| Sitagliptin | 52 | 5 | 9 | 436 | STP | WW-sources | |
| Sucralose | 127 | 31 | 966 | *43400* | Core | | B |
| Sulfadimethoxine | 0 | NA | NA | NA | | | |
| Sulfamethazine | 0 | NA | NA | NA | | | |
| Sulfamethoxazole | 81 | 6 | 24 | 678 | Core | | B |
| Sulfathiazole | 2 | 40 | 57 | 75 | Diffuse | Diffuse-general | |
| TDCPP | 54 | 5 | 32 | 204 | Core | | D |
| Temazepam | 64 | 1 | 4 | 231 | Core | | B |
| Terbutylazine | 3 | 1 | 2 | 121 | Diffuse | Diffuse-general | |
| Testosterone | 6 | 4 | 6 | 10 | Diffuse | Diffuse-specific | |
| Thiabendazole | 37 | 1 | 2 | 19 | STP | WW-sources | |
| Triamterene | 30 | 1 | 4 | 56 | STP | WW-sources | |
| Tributyl-phosphate | 55 | 3 | 29 | 204 | Core | | D |
| Triclosan | 9 | 61 | 124 | 428 | STP | STP-exclusive | |
| Trimethoprim | 106 | 1 | 4 | 531 | Core | | B |
| Trinexapac-ethyl | 11 | 5 | 29 | 128 | STP | STP-exclusive | |
| Tris(2-chloroethyl)phosphate | 37 | 5 | 9 | 260 | STP | WW-sources | |
| Valsartan | 114 | 3 | 21 | *2658* | Core | | C |
| Venlafaxine | 121 | 2 | 6 | 519 | Core | | B |
| Verapamil | 15 | 2 | 5 | 30 | STP | WW-sources | |
| Warfarin | 50 | 1 | 3 | 10 | Diffuse | Diffuse-general | |

[a] **Minimum concentration** detected above the limit of quantification.

[b] **Median and maximum concentrations** above *1000 $ng \cdot L^{-1}$* are outside of the calibration range and estimated.

**Occurrence cluster:** *Core micropollutants* were found ubiquitously; *STP micropollutants* were frequently detected in the STP outfall samples and sporadically in tributary or control site samples (*STP exclusive* = detected in STP outfall samples only; *WW-sources* = also detected sporadically in the tributary samples); *Diffuse micropollutants* were infrequently detected and were not necessarily influenced by STP outfalls (*Diffuse-specific* = only detected in specific STP outfall and tributary samples during specific sampling events; *Diffuse-general* = infrequently detected in all sample types and times).

**Normalized concentration cluster:** *Cluster A* represents agricultural micropollutants; *Clusters B and C* represent STP micropollutants which occurred at relatively high and low concentrations in the STP outfalls; *Cluster D* represents mixed-use micropollutant

## A.6 Predictors of micropollutant occurrence

**Table A10:** Results of significant ($n > 50$, $p < 0.01$, $\rho > 0.65$) Pearson correlations between summary statistics and concentrations of individual micropollutants.

| **Total detections** | | | |
|---|---|---|---|
| **Micropollutant** | **$R^2$** | **ρ** | **n** |
| Dextromethorphan | 0.73 | 0.83 | 74 |
| Naproxen | 0.74 | 0.79 | 59 |
| Carbendazim | 0.66 | 0.78 | 117 |
| Sitagliptin | 0.68 | 0.75 | 52 |
| Lidocaine | 0.63 | 0.75 | 115 |
| Losartan | 0.63 | 0.74 | 87 |
| Cotinine | 0.66 | 0.73 | 126 |
| N4-acetylsulfamethoxazole | 0.56 | 0.72 | 77 |
| Fluconazole | 0.62 | 0.71 | 82 |
| Ritalinic_Acid | 0.64 | 0.69 | 118 |
| TDCPP | 0.50 | 0.69 | 54 |
| Gabapentin | 0.61 | 0.67 | 127 |
| Temazepam | 0.61 | 0.67 | 64 |
| Pseudoephedrine | 0.63 | 0.66 | 95 |
| Levetiracetam | 0.62 | 0.66 | 105 |
| Methadone | 0.50 | 0.66 | 63 |
| Sucralose | 0.58 | 0.65 | 127 |
| **Cumulative concentrations** | | | |
| **Micropollutant** | **$R^2$** | **ρ** | **n** |
| Sitagliptin | 0.90 | 0.85 | 52 |
| Bupropion | 0.80 | 0.81 | 72 |
| Naproxen | 0.61 | 0.73 | 59 |
| Fluconazole | 0.83 | 0.70 | 82 |
| N4-acetylsulfamethoxazole | 0.78 | 0.70 | 77 |
| Dextromethorphan | 0.82 | 0.68 | 74 |
| Trimethoprim | 0.74 | 0.68 | 106 |
| Benzoylecgonine | 0.61 | 0.67 | 74 |
| Lidocaine | 0.82 | 0.66 | 115 |
| Metoprolol | 0.80 | 0.66 | 121 |
| Metformin | 0.71 | 0.66 | 120 |
| Desvenlafaxine | 0.81 | 0.65 | 122 |
| TDCPP | 0.54 | 0.65 | 54 |

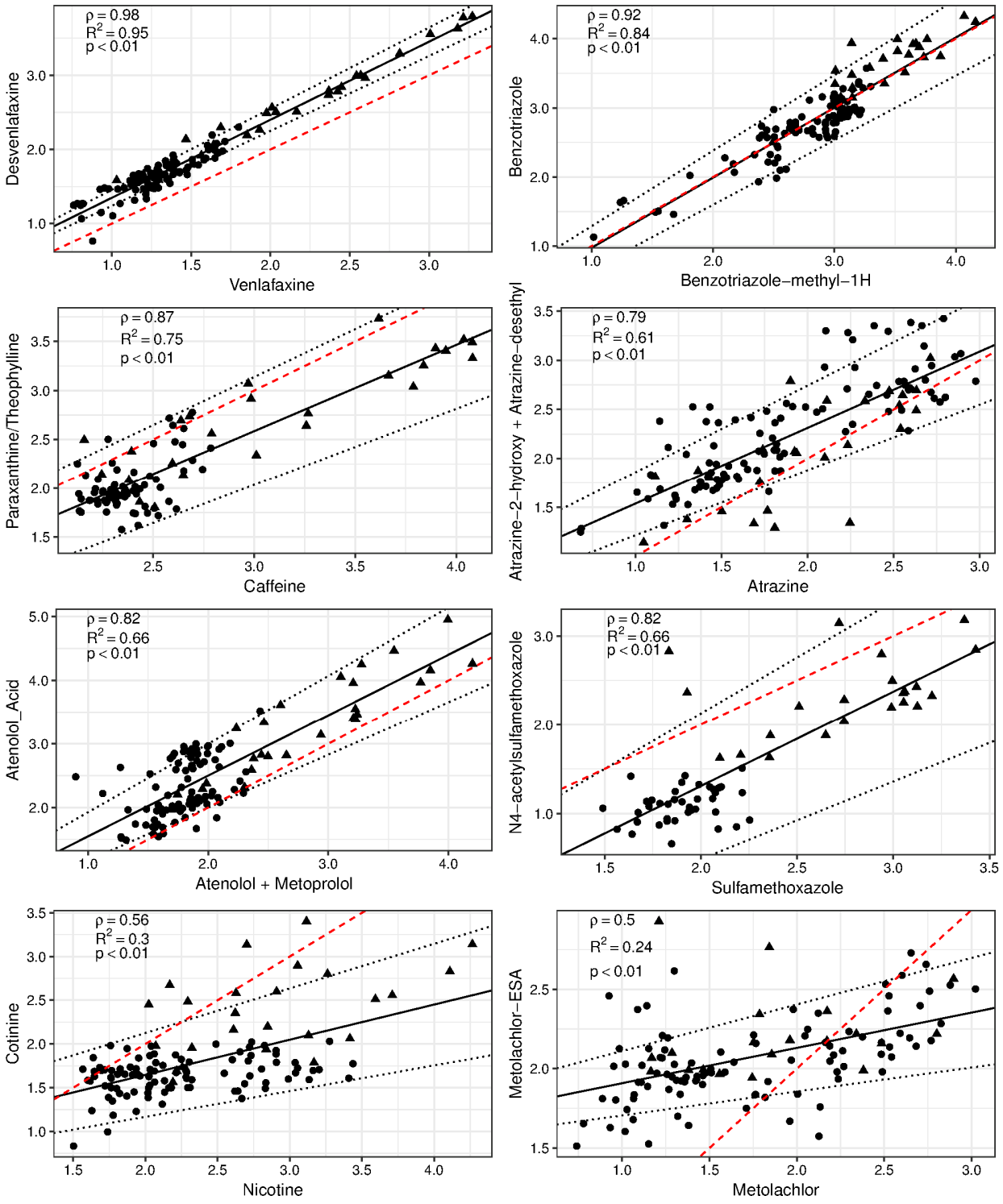



**Figure A4:** Linear regression (solid line) of parent micropollutants and transformation product pairs (log concentrations, pmol $L^{-1}$) at each sample site (circles = tributary, triangles = STP). The dotted line represents the 95% confidence interval. The red dashed line represents the 1:1 line.

# APPENDIX B – Fall Creek Monitoring Station: Highly Resolved Temporal Sampling to Prioritize the Identification of Nontarget Micropollutants in a Small Stream

## B.1 Study area

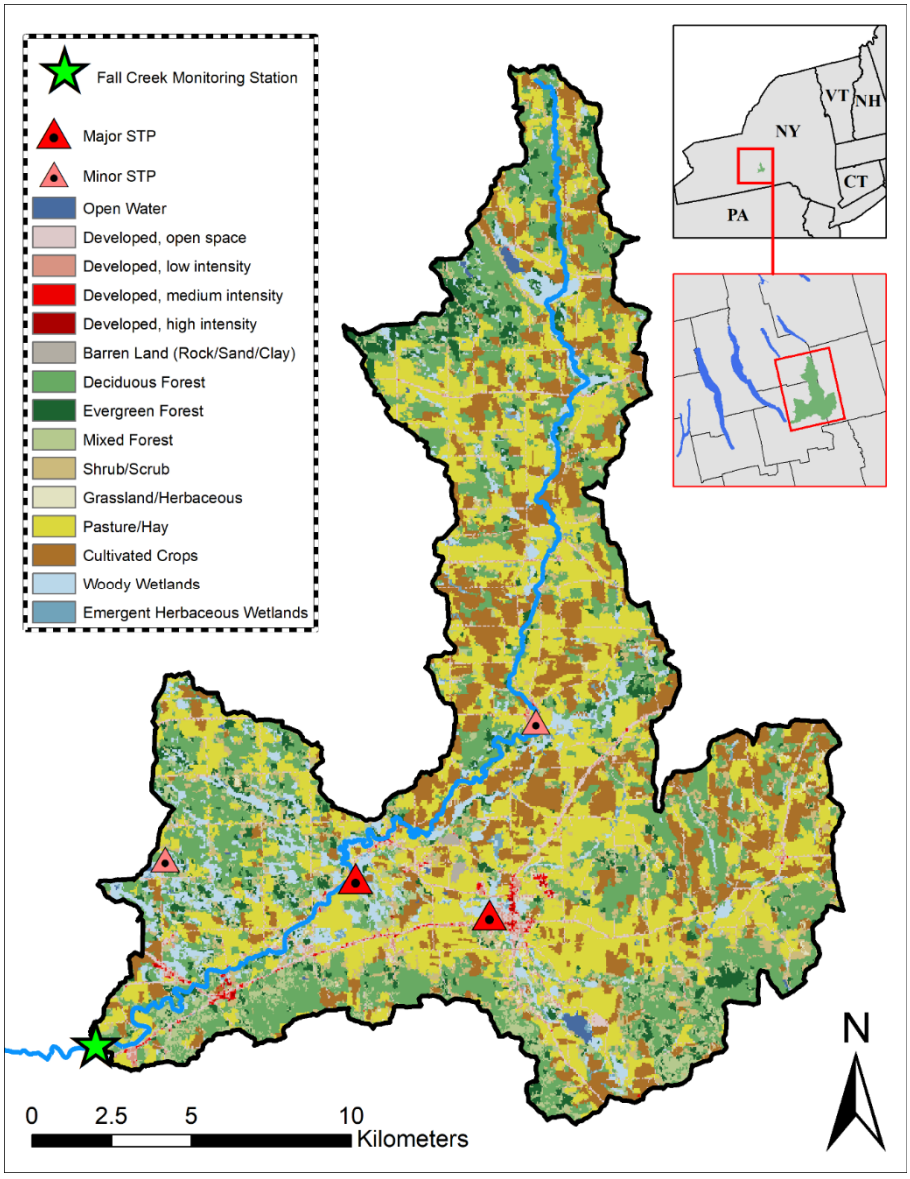



**Figure B1:** Map of the Fall Creek watershed upstream of the Fall Creek Monitoring Station (green star), locations of sewage treatment plants (red triangles), and land cover.

## B.2 Standards and reagents

**Table B1:** List of 162 target micropollutants.

| Micropollutant | CAS number | Chemical Formula | Adduct | Exact Mass (*m/z*) | RT (min) | Diagnostic Fragment (*m/z*) |
|---|---|---|---|---|---|---|
| 10,11-dihydrocarbamazepine | 3564-73-6 | C15H14N2O1 | M+H | 239.1179 | 13.9 | 194.10 |
| 1-methyl-benzotriazole | 136-85-6 | C7H7N3 | M+H | 134.0713 | 12.2 | 106.07 |
| 1-Naphthol | 90-15-3 | C10H8O1 | M+H | 145.0648 | 13.5 | 117.07 |
| 2,4-D | 94-75-7 | C8H6Cl2O3 | M-H | 218.9611 | 15.2 | 160.96 |
| 2-aminobenzimidazole | 934-32-7 | C7H7N3 | M+H | 134.0713 | 12.2 | 92.05 |
| 2-ethyl-2-phenyl-malonamide | 7206-76-0 | C11H14N2O2 | M+H | 207.1134 | 10.3 | 119.09 |
| 2-Methyl-4-chlorophenoxyacetic acid (MCPA) | 94-74-6 | C9H9Cl1O3 | M-H | 199.0167 | 15.4 | 141.01 |
| 6-benzylaminopurine | 1214-39-7 | C12H11N5 | M+H | 226.1093 | 11.6 | 148.06 |
| Abacavir | 136470-78-5 | C14H18N6O1 | M+H | 287.1620 | 10.3 | 191.10 |
| Abscisic acid | 21293-29-8 | C15H20O4 | M+H | 265.1434 | 12.9 | 187.11 |
| Acebutolol | 37517-30-9 | C18H28N2O4 | M+H | 337.2122 | 11.1 | 319.20 |
| Acephate | 30560-19-1 | C4H10N1O3P1S1 | M+H | 184.0192 | 8.6 | 142.99 |
| Acesulfame | 55589-62-3 | C4H5N1O4S1 | M-H | 161.9850 | 10.2 | 82.03 |
| Acetamiprid | 135410-20-7 | C10H11Cl1N4 | M+H | 223.0745 | 11.6 | 126.01 |
| Acetazolamide | 1424-27-7 | C4H6N4O3S2 | M+H | 222.9954 | 9.3 | 181.97 |
| Adrenosterone | 382-45-6 | C19H24O3 | M+H | 301.1803 | 13.4 | 257.15 |
| Alachlor | 15972-60-8 | C14H20Cl1N1O2 | M+H | 270.1255 | 17.2 | 148.11 |
| Albuterol | 18559-94-9 | C13H21N1O3 | M+H | 240.1599 | 9.1 | 148.08 |
| Aldicarb | 116-06-3 | C7H14N2O2S1 | M+Na | 213.0668 | 12.2 | 89.04 |
| Amcinonide | 51022-69-6 | C28H35F1O7 | M+H | 503.2445 | 18.6 | 321.15 |
| Amisulpride | 71675-85-9 | C17H27N3O4S1 | M+H | 370.1795 | 10.2 | 242.05 |
| Amitriptyline | 50-48-6 | C20H23N1 | M+H | 278.1908 | 14.5 | 233.13 |
| Amphetamine | 300-62-9 | C9H13N1 | M+H | 136.1126 | 10.1 | 91.05 |
| Ampicillin | 69-53-4 | C16H19N3O4S1 | M+H | 350.1169 | 11.1 | 192.05 |

**Table B1** (continued)

| Micropollutant | CAS number | Chemical Formula | Adduct | Exact Mass (*m/z*) | RT (min) | Diagnostic Fragment (*m/z*) |
|---|---|---|---|---|---|---|
| 10,11-dihydrocarbamazepine | 3564-73-6 | C15H14N2O1 | M+H | 239.1179 | 13.9 | 194.10 |
| 1-methyl-benzotriazole | 136-85-6 | C7H7N3 | M+H | 134.0713 | 12.2 | 106.07 |
| 1-Naphthol | 90-15-3 | C10H8O1 | M+H | 145.0648 | 13.5 | 117.07 |
| 2,4-D | 94-75-7 | C8H6Cl2O3 | M-H | 218.9611 | 15.2 | 160.96 |
| 2-aminobenzimidazole | 934-32-7 | C7H7N3 | M+H | 134.0713 | 12.2 | 92.05 |
| 2-ethyl-2-phenyl-malonamide | 7206-76-0 | C11H14N2O2 | M+H | 207.1134 | 10.3 | 119.09 |
| 2-Methyl-4-chlorophenoxyacetic acid (MCPA) | 94-74-6 | C9H9Cl1O3 | M-H | 199.0167 | 15.4 | 141.01 |
| 6-benzylaminopurine | 1214-39-7 | C12H11N5 | M+H | 226.1093 | 11.6 | 148.06 |
| Abacavir | 136470-78-5 | C14H18N6O1 | M+H | 287.1620 | 10.3 | 191.10 |
| Abscisic acid | 21293-29-8 | C15H20O4 | M+H | 265.1434 | 12.9 | 187.11 |
| Acebutolol | 37517-30-9 | C18H28N2O4 | M+H | 337.2122 | 11.1 | 319.20 |
| Acephate | 30560-19-1 | C4H10N1O3P1S1 | M+H | 184.0192 | 8.6 | 142.99 |
| Acesulfame | 55589-62-3 | C4H5N1O4S1 | M-H | 161.9850 | 10.2 | 82.03 |
| Acetamiprid | 135410-20-7 | C10H11Cl1N4 | M+H | 223.0745 | 11.6 | 126.01 |
| Acetazolamide | 1424-27-7 | C4H6N4O3S2 | M+H | 222.9954 | 9.3 | 181.97 |
| Adrenosterone | 382-45-6 | C19H24O3 | M+H | 301.1803 | 13.4 | 257.15 |
| Alachlor | 15972-60-8 | C14H20Cl1N1O2 | M+H | 270.1255 | 17.2 | 148.11 |
| Albuterol | 18559-94-9 | C13H21N1O3 | M+H | 240.1599 | 9.1 | 148.08 |
| Aldicarb | 116-06-3 | C7H14N2O2S1 | M+Na | 213.0668 | 12.2 | 89.04 |
| Amcinonide | 51022-69-6 | C28H35F1O7 | M+H | 503.2445 | 18.6 | 321.15 |
| Amisulpride | 71675-85-9 | C17H27N3O4S1 | M+H | 370.1795 | 10.2 | 242.05 |
| Amitriptyline | 50-48-6 | C20H23N1 | M+H | 278.1908 | 14.5 | 233.13 |
| Amphetamine | 300-62-9 | C9H13N1 | M+H | 136.1126 | 10.1 | 91.05 |

**Table B1** (continued)

| Micropollutant | CAS number | Chemical Formula | Adduct | Exact Mass (*m/z*) | RT (min) | Diagnostic Fragment (*m/z*) |
|---|---|---|---|---|---|---|
| Ampicillin | 69-53-4 | C16H19N3O4S1 | M+H | 350.1169 | 11.1 | 192.05 |
| Atenolol | 29122-68-7 | C14H22N2O3 | M+H | 267.1703 | 9.1 | 190.09 |
| Atenolol acid | 56392-14-4 | C14H21N1O4 | M+H | 268.1549 | 10.2 | 191.07 |
| Atomoxetine | 83015-26-3 | C17H21N1O1 | M+H | 256.1696 | 13.8 | 224.08 |
| Atorvastatin | 134523-00-5 | C33H35F1N2O5 | M+H | 559.2608 | 17.8 | 440.22 |
| Atrazine | 1912-24-9 | C8H14Cl1N5 | M+H | 216.1010 | 14.2 | 174.05 |
| Atrazine-2-hydroxy | 2163-68-0 | C8H15N5O1 | M+H | 198.1349 | 10.8 | 156.09 |
| Atrazine-desethyl | 6190-65-4 | C6H10Cl1N5 | M+H | 188.0697 | 11.9 | 146.02 |
| Atropine | 51-55-8 | C17H23N1O3 | M+H | 290.1751 | 10.7 | 260.16 |
| Azoxystrobin | 131860-33-8 | C22H17N3O5 | M+H | 404.1241 | 15.6 | 372.10 |
| Baclofen | 1134-47-0 | C10H12Cl1N1O2 | M+H | 214.0635 | 10.4 | 151.03 |
| Bendiocarb | 22781-23-3 | C11H13N1O4 | M+H | 224.0917 | 13.1 | 167.07 |
| Bentazon | 25057-89-0 | C10H12N2O3S1 | M-H | 239.0485 | 13.7 | 175.09 |
| Benzisothiazolin-3-one (BIT) | 2634-33-5 | C7H5N1O1S1 | M+H | 152.0164 | 11.5 | 109.01 |
| Benzophenone | 119-61-9 | C13H10O1 | M+H | 183.0804 | 15.9 | 105.03 |
| Benzotriazole | 95-14-7 | C6H5N3 | M+H | 120.0556 | 11.0 | 92.05 |
| Benzoylecgonine | 519-09-5 | C16H19N1O4 | M+H | 290.1387 | 11.0 | 168.10 |
| Bromacil | 314-40-9 | C9H13Br1N2O2 | M+H | 261.0233 | 13.0 | 204.96 |
| Bupropion | 34841-39-9 | C13H18Cl1N1O1 | M+H | 240.1155 | 12.0 | 166.04 |
| Butalbital | 77-26-9 | C11H16N2O3 | M-H | 223.1083 | 13.0 | 136.37 |
| Caffeine | 58-08-2 | C8H10N4O2 | M+H | 195.0876 | 10.7 | 138.07 |
| Candesartan | 139481-59-7 | C24H20N6O3 | M+H | 441.1670 | 15.3 | 263.13 |
| Carbamazepine | 298-46-4 | C15H12N2O1 | M+H | 237.1022 | 13.5 | 194.10 |
| Carbamazepine-10,11-epoxide | 36507-30-9 | C15H12N2O2 | M+H | 253.0972 | 12.2 | 180.08 |

**Table B1** (continued)

| Micropollutant | CAS number | Chemical Formula | Adduct | Exact Mass (*m/z*) | RT (min) | Diagnostic Fragment (*m/z*) |
|---|---|---|---|---|---|---|
| Carbaryl | 63-25-2 | C12H11N1O2 | M+H | 202.0873 | 13.5 | 145.06 |
| Carbendazim | 10605-21-7 | C9H9N3O2 | M+H | 192.0768 | 10.1 | 160.05 |
| Carbofuran | 1563-66-2 | C12H15N1O3 | M+H | 222.1135 | 13.0 | 165.09 |
| Carisoprodol | 78-44-4 | C12H24N2O4 | M+H | 261.1814 | 14.4 | 158.12 |
| Celecoxib | 169590-42-5 | C17H14F3N3O2S1 | M+H | 382.0837 | 17.7 | 362.08 |
| Chloridazon | 1698-60-8 | C10H8Cl1N3O1 | M+H | 222.0428 | 11.6 | 128.99 |
| Chlorpheniramine | 132-22-9 | C16H19Cl1N2 | M+H | 275.1315 | 11.9 | 230.07 |
| Cimetidine | 51481-61-9 | C10H16N6S1 | M+H | 253.1235 | 9.2 | 159.07 |
| Citalopram | 59729-33-8 | C20H21F1N2O1 | M+H | 325.1716 | 12.6 | 262.10 |
| Clarithromycin | 81103-11-9 | C38H69N1O13 | M+H | 748.4842 | 15.4 | 590.39 |
| Clofibric acid | 882-09-7 | C10H11Cl1O3 | M-H | 213.0313 | 15.7 | 126.99 |
| Codeine | 76-57-3 | C18H21N1O3 | M+H | 300.1599 | 9.4 | 215.11 |
| Corticosterone | 50-22-6 | C21H30O4 | M+H | 347.2217 | 14.8 | 329.21 |
| Coumarin | 91-64-5 | C9H6O2 | M+H | 147.0441 | 12.1 | 91.05 |
| Cyanazine | 21725-46-2 | C9H13Cl1N6 | M+H | 241.0963 | 12.6 | 214.09 |
| Cyflufenamid | 180409-60-3 | C20H17F5N2O2 | M+H | 413.1288 | 19.0 | 241.04 |
| DEET | 134-62-3 | C12H17N1O1 | M+H | 192.1383 | 14.3 | 119.05 |
| Dehydroacetic acid | 520-45-6 | C8H8O4 | M+H | 169.0495 | 12.1 | 85.03 |
| Desvenlafaxine | 93413-62-8 | C16H25N1O2 | M+H | 264.1958 | 11.1 | 201.13 |
| Dexamethasone | 50-02-2 | C22H29F1O5 | M+H | 393.2077 | 14.8 | 237.13 |
| Dextromethorphan | 125-71-3 | C18H25N1O1 | M+H | 272.2014 | 12.6 | 147.08 |
| Diazepam | 439-14-5 | C16H13Cl1N2O1 | M+H | 285.0789 | 15.7 | 154.04 |
| Diazinon | 333-41-5 | C12H21N2O3P1S1 | M+H | 305.1083 | 18.4 | 169.08 |
| Diclofenac | 15307-86-5 | C14H11Cl2N1O2 | M+H | 296.0239 | 18.3 | 215.05 |

**Table B1** (continued)

| Micropollutant | CAS number | Chemical Formula | Adduct | Exact Mass (*m/z*) | RT (min) | Diagnostic Fragment (*m/z*) |
|---|---|---|---|---|---|---|
| Diethyl phthalate | 84-66-2 | C12H14O4 | M+H | 223.0965 | 14.6 | 149.02 |
| Diltiazem | 42399-41-7 | C22H26N2O4S1 | M+H | 415.1686 | 13.1 | 178.03 |
| Dimethachlor | 50563-36-5 | C13H18Cl1N1O2 | M+H | 256.1099 | 14.8 | 224.08 |
| Dimethoate | 60-51-5 | C5H12N1O3P1S2 | M+H | 230.0069 | 11.6 | 142.99 |
| Diphenhydramine | 58-73-1 | C17H21N1O1 | M+H | 256.1701 | 12.8 | 167.09 |
| Diuron | 330-54-1 | C9H10Cl2N2O1 | M+H | 233.0243 | 14.8 | 159.97 |
| Efavirenz | 154598-52-4 | C14H9Cl1F3N1O2 | M+H | 316.0352 | 18.3 | 244.01 |
| Estriol | 50-27-1 | C18H24O3 | M+Na | 311.1618 | 12.8 | 226.89 |
| Estrone | 53-16-7 | C18H22O2 | M+H | 271.1698 | 15.9 | 157.06 |
| Ethofumesate | 26225-79-6 | C13H18O5S1 | M+H | 287.0947 | 15.7 | 161.06 |
| Ethyl butylacetylaminopropionate | 52304-36-6 | C11H21N1O3 | M+H | 216.1599 | 13.5 | 86.10 |
| Famciclovir | 104227-87-4 | C14H19N5O4 | M+H | 322.1510 | 10.5 | 202.11 |
| Famotidine | 76824-35-6 | C8H15N7O2S3 | M+H | 338.0527 | 9.2 | 189.03 |
| Fexofenadine | 83799-24-0 | C32H39N1O4 | M+H | 502.2957 | 14.0 | 484.28 |
| Fluconazole | 86386-73-4 | C13H12F2N6O1 | M+H | 307.1113 | 11.5 | 220.07 |
| Flucytosine | 2022-85-7 | C4H4F1N3O1 | M+H | 130.0411 | 9.1 | 71.06 |
| Fluoxetine | 54910-89-3 | C17H18F3N1O1 | M+H | 310.1413 | 15.0 | 265.16 |
| Folic acid | 59-30-3 | C19H19N7O6 | M+H | 442.1470 | 10.7 | 295.09 |
| Furosemid | 54-31-9 | C12H11Cl1N2O5S1 | M-H | 328.9999 | 12.7 | 204.98 |
| Gabapentin | 60142-96-3 | C9H17N1O2 | M+H | 172.1332 | 10.0 | 137.10 |
| Gemfibrozil | 25812-30-0 | C15H22O3 | M+H | 251.1641 | 20.2 | 129.09 |
| Gibberellic Acid | 77-06-5 | C19H22O6 | M-H | 345.1333 | 11.5 | 221.13 |
| Hexamethylphosphoramide | 680-31-9 | C6H18N3O1P1 | M+H | 180.1260 | 11.6 | 135.07 |
| Hexazinone | 51235-04-2 | C12H20N4O2 | M+H | 253.1659 | 13.0 | 171.09 |

**Table B1** (continued)

| Micropollutant | CAS number | Chemical Formula | Adduct | Exact Mass (*m/z*) | RT (min) | Diagnostic Fragment (*m/z*) |
|---|---|---|---|---|---|---|
| Hydrocodone | 125-29-1 | C18H21N1O3 | M+H | 300.1599 | 9.9 | 199.08 |
| Hydrocortisone | 50-23-7 | C21H30O5 | M+H | 363.2171 | 13.8 | 327.19 |
| Ibuprofen | 15687-27-1 | C13H18O2 | M+Na | 229.1199 | 18.6 | 181.99 |
| Imidacloprid | 138261-41-3 | C9H10Cl1N5O2 | M+H | 256.0601 | 11.1 | 175.10 |
| Iodocarb | 55406-53-6 | C8H12I1N1O2 | M+H | 281.9985 | 14.2 | 164.92 |
| Ioxynil | 1689-83-4 | C7H3I2N1O1 | M-H | 369.8221 | 15.3 | 230.92 |
| Irbesartan | 138402-11-6 | C25H28N6O1 | M+H | 429.2397 | 14.7 | 386.22 |
| Isophorone Diisocyanate | 4098-71-9 | C12H18O2N2 | M+K | 261.0995 | 13.9 | 204.96 |
| Isoproturon | 34123-59-6 | C12H18N2O1 | M+H | 207.1492 | 14.5 | 134.10 |
| Ketamine | 6740-88-1 | C13H16Cl1N1O1 | M+H | 238.0998 | 10.9 | 179.06 |
| Ketoprofen | 22071-15-4 | C16H14O3 | M+H | 255.1015 | 15.3 | 209.10 |
| Lamotrigine | 84057-84-1 | C9H7Cl2N5 | M+H | 256.0151 | 11.6 | 210.98 |
| Lidocaine | 137-58-6 | C14H22N2O1 | M+H | 235.1810 | 10.6 | 86.10 |
| Linuron | 330-55-2 | C9H10Cl2N2O2 | M+H | 249.0192 | 15.8 | 159.97 |
| Losartan | 114798-26-4 | C22H23Cl1N6O1 | M+H | 423.1695 | 14.7 | 235.10 |
| Mabuterol | 56341-08-3 | C13H18Cl1F3N2O1 | M+H | 311.1138 | 11.6 | 237.04 |
| Malaoxon | 1634-78-2 | C10H19O7P1S1 | M+H | 315.0667 | 13.1 | 255.07 |
| Mecoprop | 93-65-2 | C10H11Cl1O3 | M-H | 213.0313 | 16.6 | 141.01 |
| Meprobamate | 57-53-4 | C9H18N2O4 | M+H | 219.1345 | 12.3 | 203.14 |
| Metalaxyl | 57837-19-1 | C15H21N1O4 | M+H | 280.1543 | 14.4 | 160.11 |
| Metamitron | 41394-05-2 | C10H10N4O1 | M+H | 203.0938 | 11.5 | 175.10 |
| Metaxalone | 1665-48-1 | C12H15N1O3 | M+H | 222.1135 | 14.0 | 133.10 |
| Methocarbamol | 532-03-6 | C11H15N1O5 | M+H | 242.1028 | 11.7 | 163.08 |
| Methomyl | 16752-77-5 | C5H10N2O2S1 | M+H | 163.0541 | 10.2 | 102.97 |

**Table B1** (continued)

| Micropollutant | CAS number | Chemical Formula | Adduct | Exact Mass (*m/z*) | RT (min) | Diagnostic Fragment (*m/z*) |
|---|---|---|---|---|---|---|
| Metolachlor | 51218-45-2 | C15H22Cl1N1O2 | M+H | 284.1412 | 17.2 | 252.12 |
| Metolachlor-ESA | 171118-09-5 | C15H23N1O5S1 | M+H | 330.1370 | 14.7 | 298.11 |
| Metoprolol | 37350-58-6 | C15H25N1O3 | M+H | 268.1907 | 11.2 | 159.08 |
| Metribuzin | 21087-64-9 | C8H14N4O1S1 | M+H | 215.0961 | 12.9 | 187.10 |
| Metsulfuron-methyl | 74223-64-6 | C14H15N5O6S1 | M+H | 382.0816 | 13.2 | 340.12 |
| Molinate | 2212-67-1 | C9H17N1O1S1 | M+H | 188.1109 | 16.2 | 126.09 |
| N,N-didesmethyl venlafaxine | 93413-77-5 | C15H23N1O2 | M+H | 250.1802 | 12.3 | 147.08 |
| N4-acetylsulfamethoxazole | 21312-10-7 | C12H13N3O4S1 | M+H | 296.0700 | 11.7 | 198.02 |
| Nadolol | 42200-33-9 | C17H27N1O4 | M+H | 310.2013 | 10.4 | 254.14 |
| Naproxen | 22204-53-1 | C14H14O3 | M+H | 231.1015 | 15.8 | 185.10 |
| Oxcarbazepine | 28721-07-5 | C15H12N2O2 | M+H | 253.0977 | 12.6 | 208.07 |
| Oxybenzone | 131-57-7 | C14H12O3 | M+H | 229.0864 | 17.7 | 151.04 |
| Paraxanthine | 611-59-6 | C7H8N4O2 | M+H | 181.0720 | 9.9 | 124.05 |
| Pentoxifylline | 6493-05-6 | C13H18N4O3 | M+H | 279.1457 | 11.7 | 181.07 |
| Perfluorobutanoic Acid (PFBA) | 375-22-4 | C4H1F7O2 | M-H | 212.9792 | 12.5 | 168.99 |
| Perfluorooctanoic Acid (PFOA) | 335-67-1 | C8H1F15O2 | M-H | 412.9665 | 19.1 | 218.99 |
| Phenobarbital | 50-06-6 | C12H12N2O3 | M-H | 231.0770 | 12.2 | 188.07 |
| Phenytoin | 57-41-0 | C15H12N2O2 | M+H | 253.0977 | 13.3 | 182.10 |
| Pirimicarb | 23103-98-2 | C11H18N4O2 | M+H | 239.1508 | 11.1 | 182.13 |
| Pirimiphos-ethyl | 23505-41-1 | C13H24N3O3P1S1 | M+H | 334.1349 | 19.7 | 198.11 |
| Primidone | 125-33-7 | C12H14N2O2 | M+H | 219.1128 | 11.6 | 162.09 |
| Progesterone | 57-83-0 | C21H30O2 | M+H | 315.2324 | 18.3 | 297.22 |
| Prohexadione | 127277-53-6 | C10H12O5 | M-H | 211.0606 | 12.5 | 125.02 |
| Prometon | 1610-18-0 | C10H19N5O1 | M+H | 226.1662 | 12.6 | 184.12 |

**Table B1** (continued)

| Micropollutant | CAS number | Chemical Formula | Adduct | Exact Mass (*m/z*) | RT (min) | Diagnostic Fragment (*m/z*) |
|---|---|---|---|---|---|---|
| Propachlor | 1918-16-7 | C11H14Cl1N1O1 | M+H | 212.0836 | 14.3 | 170.04 |
| Propachlor-ESA | 123732-85-4 | C11H15N1O4S1 | M-H | 256.0638 | 12.8 | 162.02 |
| Propachlor-OXA | 70628-36-3 | C11H13N1O3 | M+H | 208.0968 | 12.6 | 120.04 |
| Propazine | 139-40-2 | C9H16Cl1N5 | M+H | 230.1167 | 15.5 | 146.02 |
| Propoxur | 114-26-1 | C11H15N1O3 | M+H | 210.1125 | 12.9 | 111.04 |
| Propranolol | 525-66-6 | C16H21N1O2 | M+H | 260.1651 | 12.6 | 183.08 |
| Propyzamide | 23950-58-5 | C12H11Cl2N1O1 | M+H | 256.0296 | 16.3 | 172.96 |
| Pseudoephedrine | 299-42-3 | C10H15N1O1 | M+H | 166.1229 | 9.6 | 133.09 |
| Pyrazophos | 13457-18-6 | C14H20N3O5P1S1 | M+H | 374.0934 | 18.9 | 222.09 |
| Ranitidine | 66357-35-5 | C13H22N4O3S1 | M+H | 315.1485 | 9.1 | 176.05 |
| Ritalinic acid | 19395-41-6 | C13H17N1O2 | M+H | 220.1332 | 10.9 | 84.08 |
| Sertraline | 79617-96-2 | C17H17Cl2N1 | M+H | 306.0816 | 15.5 | 275.04 |
| Siduron | 1982-49-6 | C14H20N2O1 | M+H | 233.1654 | 15.7 | 137.07 |
| Simazine | 122-34-9 | C7H12Cl1N5 | M+H | 202.0854 | 12.9 | 132.03 |
| Sitagliptin | 486460-32-6 | C16H15F6N5O1 | M+H | 408.1259 | 11.4 | 235.08 |
| Sucralose | 56038-13-2 | C12H19Cl3O8 | M+FA-H | 447.0499 | 10.8 | 278.15 |
| Sulfadimethoxine | 122-11-2 | C12H14N4O4S1 | M+H | 311.0808 | 11.9 | 237.04 |
| Sulfamethazine | 57-68-1 | C12H14N4O2S1 | M+H | 279.0910 | 10.4 | 204.04 |
| Sulfamethoxazole | 723-46-6 | C10H11N3O3S1 | M+H | 254.0594 | 11.0 | 156.01 |
| Sulfanilic acid | 121-57-3 | C6H7N1O3S1 | M-H | 174.0116 | 6.5 | 79.96 |
| Sulfathiazole | 72-14-0 | C9H9N3O2S2 | M+H | 256.0209 | 9.6 | 156.01 |
| Temazepam | 846-50-4 | C16H13Cl1N2O2 | M+H | 301.0744 | 14.9 | 255.07 |
| Terbutylazine | 5915-41-3 | C9H16Cl1N5 | M+H | 230.1167 | 15.8 | 174.05 |
| Testosterone | 58-22-0 | C19H28O2 | M+H | 289.2168 | 16.3 | 253.20 |

**Table B1** (continued)

| Micropollutant | CAS number | Chemical Formula | Adduct | Exact Mass (*m/z*) | RT (min) | Diagnostic Fragment (*m/z*) |
|---|---|---|---|---|---|---|
| Thiabendazole | 148-79-8 | C10H7N3S1 | M+H | 202.0433 | 10.6 | 175.03 |
| Triamterene | 396-01-0 | C12H11N7 | M+H | 254.1154 | 10.9 | 237.09 |
| Tributyl-phosphate | 126-73-8 | C12H27O4P1 | M+H | 267.1725 | 19.8 | 155.05 |
| Triclosan | 3380-34-5 | C12H7Cl3O2 | M-H | 286.9428 | 20.3 | 165.89 |
| Trimethoprim | 738-70-5 | C14H18N4O3 | M+H | 291.1451 | 10.2 | 245.10 |
| Trinexapac-ethyl | 95266-40-3 | C13H16O5 | M+H | 253.1070 | 14.7 | 183.03 |
| Tris(1,3-dichloro-2-propyl)phosphate (TDCPP) | 13674-87-8 | C9H15Cl6O4P1 | M+H | 428.8912 | 18.6 | 98.98 |
| Tris(2-carboxyethyl)phosphine (TCEP) | 51805-45-9 | C6H12Cl3O4P1 | M+H | 251.0679 | 13.2 | 159.02 |
| Valsartan | 137862-53-4 | C24H29N5O3 | M+H | 436.2348 | 16.6 | 235.10 |
| Venlafaxine | 93413-69-5 | C17H27N1O2 | M+H | 278.2114 | 12.2 | 121.06 |
| Verapamil | 52-53-9 | C27H38N2O4 | M+H | 455.2910 | 13.2 | 303.21 |
| Warfarin | 2610-86-8 | C19H16O4 | M+H | 309.1121 | 16.2 | 251.07 |

**Table B2:** List of 33 isotope labeled internal standards.

| Micropollutant | CAS number | Chemical Formula | Adduct | RT (min) |
|---|---|---|---|---|
| Atenolol-d7 | 1202864-50-3 | C14D7H15N2O3 | M+H | 9.1 |
| Atrazine-d5 | 163165-75-1 | C8H9D5Cl1N5 | M+H | 14.1 |
| Azoxystrobin-d4 | 1346606-39-0 | C22H13D4N3O5 | M+H | 15.5 |
| Caffeine-13C3 | n/a | [13]C3C5H10N4O2 | M+H | 10.6 |
| Carbamazepine-13C6 | n/a | [13]C6C9H12N2O1 | M+H | 13.5 |
| Carbaryl-d7 | 362049-56-7 | C12H4D7N1O2 | M+H | 13.4 |
| Carbofuran-d3 | 1007459-98-4 | C12H12D3N1O3 | M+H | 13.0 |
| Celecoxib-d4 | 544686-20-6 | C17H10D4F3N3O2S1 | M+H | 17.7 |
| Citalopram-d6 | 1246819-94-2 | C20D6H15F1N2O1 | M+H | 12.6 |
| Diazinon-d10 | 100155-47-3 | C12D10H11N2O3P1S1 | M+H | 18.4 |
| Dimethoate-d6 | 1219794-81-6 | C5H6D6N1O3P1S2 | M+H | 11.6 |
| Estrone-d2 | 350820-16-5 | C18H20D2O2 | M+H | 15.9 |
| Fexofenadine-d6 | 548783-71-7 | C32H33D6N1O4 | M+H | 14.1 |
| Fluoxetine-d5 | 1173020-43-3 | C17D5H13F3N1O1 | M+H | 15.0 |
| Gembifrozil-d6 | 1184986-45-5 | C15H16D6O3 | M+H | 20.2 |
| Ibuprofen-d3 | 121662-14-4 | C13D3H15O2 | M+Na | 18.6 |
| Imidacloprid-d4 | 1015855-75-0 | C9D4H6Cl1N5O2 | M+H | 11.1 |
| Iodocarb-d9 | 1246815-08-6 | C8H3D9I1N1O2 | M+H | 14.2 |
| Isoproturon-d6 | 217487-17-7 | C12H12D6N2O1 | M+H | 14.4 |
| Mecoprop-d3 | 352431-15-3 | C10H8D3Cl1O3 | M-H | 16.6 |
| Metoprolol Acid-d5 | 1215404-47-9 | C14H16D5N1O4 | M+H | 10.2 |
| Naproxen-methoxy-d3 | 958293-79-3 | C14D3H11O3 | M+H | 15.7 |
| Oxybenzone-d3 | n/a | C14H9D3O3 | M+H | 17.6 |
| Pirimicarb-d6 | 1015854-66-6 | C11D6H12N4O2 | M+H | 11.0 |
| rac-Efavirenz-d4 | 1246812-58-7 | C14H5D4Cl1F3N1O2 | M+H | 18.3 |
| Ranitidine-d6 | 1185238-09-8 | C13H16D6N4O3S1 | M+H | 9.1 |
| Sucralose-d6 | 1459161-55-7 | C12H13D6Cl3O8 | M+FA-H | 10.8 |
| Sulfadimethoxine-d6 | 73068-02-7 | C12D6H8N4O4S1 | M+H | 11.9 |
| Sulfamethoxazole-phenyl-13C6 | 1196157-90-0 | [13]C6C4H11N3O3S1 | M+H | 11.0 |
| Thiabendazole-d4 | 1190007-20-5 | C10H3D4N3S1 | M+H | 10.6 |
| Triclosan-d3 | 1020719-98-5 | C12H4D3Cl3O2 | M-H | 20.8 |
| Trimethoprim-d9 | 1189460-62-5 | C14D9H9N4O3 | M+H | 10.2 |
| Tris-2-chloroethyl-phosphate-d12 | 1276500-47-0 | C6D12Cl3O4P1 | M+H | 13.2 |
| Venlafaxine-d6 | 1062606-12-5 | C17D6H21N1O2 | M+H | 12.2 |

## B.3 Analytical method

**Table B3:** Loading and elution pump gradients.

| **Retention time** (min) | **Flow** ($mL \cdot min^{-1}$) | **%A** (Water) | **%B** (Methanol) |
|---|---|---|---|
| **Loading pump** | | | |
| 0.0 | 1.000 | 99 | 1 |
| 5.1 | 1.000 | 99 | 1 |
| 5.11 | 0.000 | 99 | 1 |
| 30.2 | 0.000 | 99 | 1 |
| 30.3 | 1.000 | 2 | 98 |
| 34.3 | 1.000 | 2 | 98 |
| 34.4 | 1.000 | 99 | 1 |
| 35.6 | 1.000 | 99 | 1 |
| **Elution pump** | | | |
| 0.0 | 0.200 | 95 | 5 |
| 6.1 | 0.200 | 95 | 5 |
| 10.6 | 0.200 | 50 | 50 |
| 23.6 | 0.200 | 5 | 95 |
| 30.1 | 0.200 | 5 | 95 |
| 30.2 | 0.200 | 95 | 5 |
| 35.6 | 0.200 | 95 | 5 |

A = LC-MS-grade water with 0.1% formic acid. B = LC-MS-grade methanol with 0.1% formic acid.

**Table B4:** Source, full-scan MS, and dd-$MS^2$ parameters.

| **Source parameters** | |
|---|---|
| Voltage | +4.0 kV |
| Sheath gas flow | 40 AU |
| Auxiliary gas flow | 20 AU |
| Capillary temperature | 320 °C |
| S-lens RF level | 50 AU |
| Auxiliary gas heater temperature | 50 °C |
| **Full-scan MS** | |
| *m/z* range | 100-1000 |
| Resolution | 140,000 at 200 *m/z* |
| automatic gain control target | 500,000 |
| Maximum injection time | 200 ms |
| **dd-$MS^2$** | |
| Resolution | 14,500 at 200 *m/z* |
| Automatic gain control target | 200,000 |
| Maximum injection time | 100 ms |
| Loop count | 3 |
| Isolation window | 1 *m/z* |
| Underfill ratio | 0.1% |
| Dynamic exclusion | 8 s |
| When idle | Pick others |

## B.4 Data analysis workflow

**Table B5:** EnviMass parameters.

| | |
|---|---|
| **Peak picking** | |
| Maximum RT gap in an EIC | 300 s |
| Maximum *m/z* deviation of an EIC centroid | 5 ppm |
| Minimum number of points per peak | 5 within 7.5 s |
| Maximum RT gap to be interpolated | 2.5 s |
| Maximum RT width of a peak | 30 s |
| Minimum intensity threshold | $10^4$ |
| Minimum signal/noise | 5 |
| Minimum signal/base | 2 |
| Maximum possible number of peaks with EIC | 3 |
| **Mass recalibration** | |
| Reference compounds | Isotope labeled internal standards (ILIS) |
| Maximum allowable *m/z* correction | 10 ppm |
| Maximum allowable RT difference | 30 s |
| **Replicates** | |
| *m/z* tolerance | 5 ppm |
| RT tolerance | 30 s |
| **Screening (target and ILIS)** | |
| RT tolerance from expected | 30 s |
| RT tolerance within isotope pattern | 10 s |
| *m/z* tolerance | 5 ppm |
| Intensity tolerance | 30% |
| **Intensity normalization based on ILIS profiles** | |
| Minimum files covered by each ILIS profile | 90% |
| Screening threshold | 0.8 |
| Minimum number of ILIS profile peaks | 15 |
| Number of profiles in subsampling | 100 |
| **Profiling** | |
| *m/z* tolerance | 3 ppm |
| RT tolerance | 30 s |
| **Blind** | |
| *m/z* tolerance | 3 ppm |
| RT tolerance | 15 s |
| **Componentization (isotopologue/adduct)** | |
| *m/z* tolerance | 2.5 ppm |
| RT tolerance | 5 s |
| Adducts (Positive) | $[M+H]^+$, $[M+Na]^+$, $[M+K]^+$, $[M]^+$ |
| Adducts (Negative) | $[M-H]^-$, $[M+FA-H]^-$, $[M+Cl]^-$, $[M]^-$ |

**Table B6:** Identification confidence levels for high resolution mass spectrometry analyses adapted from Schymanski et al., 2014.[17]

| Confirmation Level | | Confirmation method |
|---|---|---|
| Level 1 | Confirmed structure | Matching MS, $MS^2$, and RT (reference standard) |
| Level 1L | Confirmed structure | Matching MS, $MS^2$, and RT (in-house library) |
| Level 2L | Probable structure | MS and matching library $MS^2$ (MoNA & mzcloud) |
| Level 2P | Probable structure | MS and matching predicted $MS^2$ (MetFrag) |
| Level 3 | Tentative candidate(s) | MS, $MS^2$ evidence, and sampling data |
| Level 4 | Unequivocal molecular formula | Distinct MS isotopic signature |
| Level 5 | Exact mass of interest | Prioritized profile |

Confirmation levels are accumulative and sequential.

### B.5 MS feature profiles of target micropollutants

Temporal profiles for the 31 target micropollutants (excluding desvenlafaxine and atrazine which are provided in **Figure 3.1**) that were identified in this study are provided in **Figures B2 – S30** and are organized alphabetically. Certain micropollutant profiles appear to have scattered non-detects across the study period (*e.g.* **Figure B16**), which is most frequently due to the failure of the peak picking algorithm to capture poor peak shapes in certain samples. Other micropollutant profiles appear to have periods of non-detects with otherwise nearly consistent normalized intensities (*e.g.* **Figure B3**), which is likely due to their abundance near the limit of detection. However, we note that despite these limitations, micropollutants with similar temporal profiles still clustered closely together (*e.g.* lamotrigine & losartan, and atenolol acid & ritalinic acid).

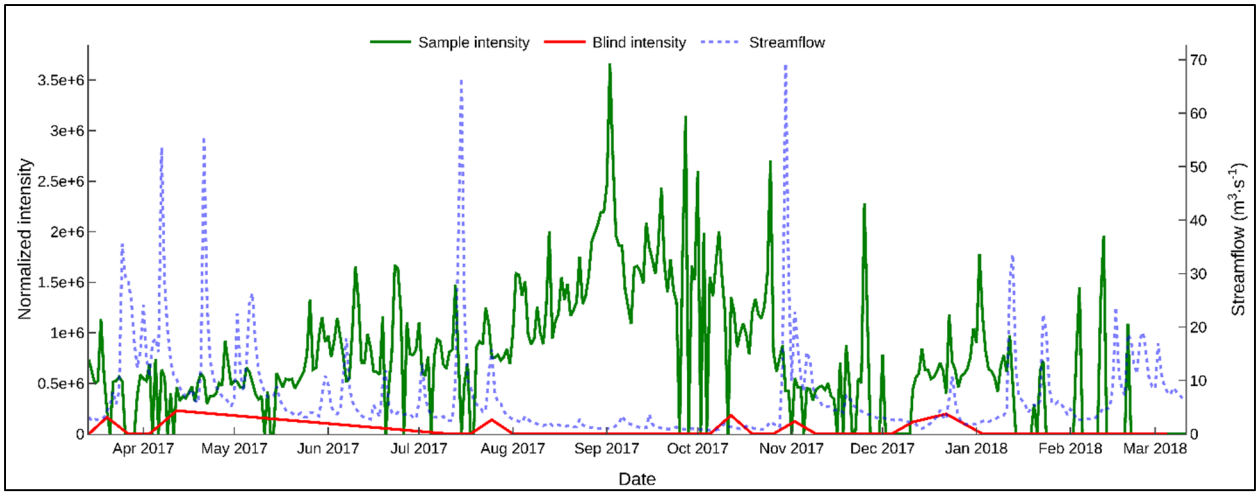



**Figure B2:** Temporal profile of 1-methyl-benzotriazole.

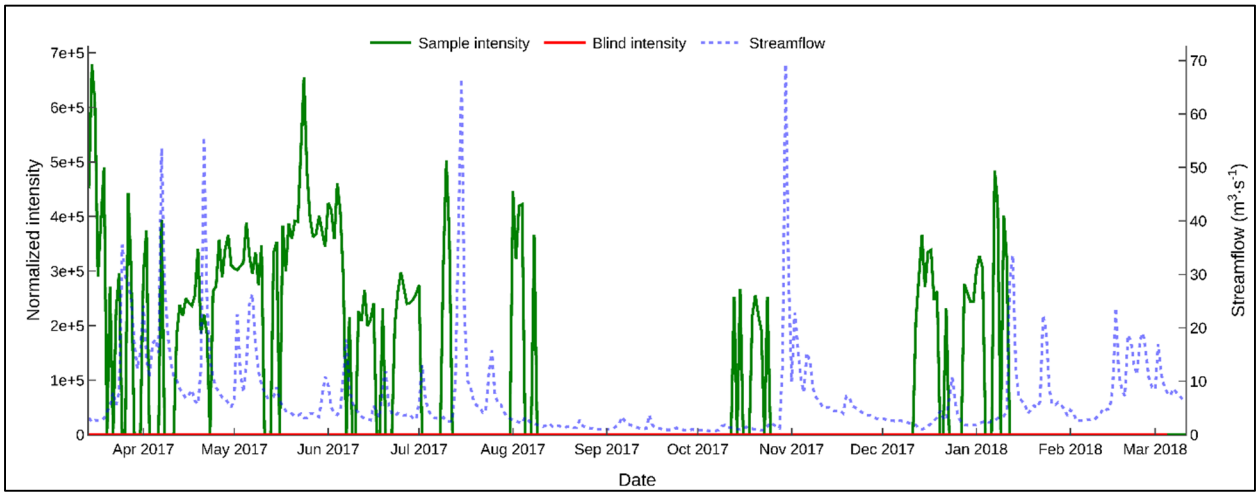



**Figure B3:** Temporal profile of atenolol-acid.

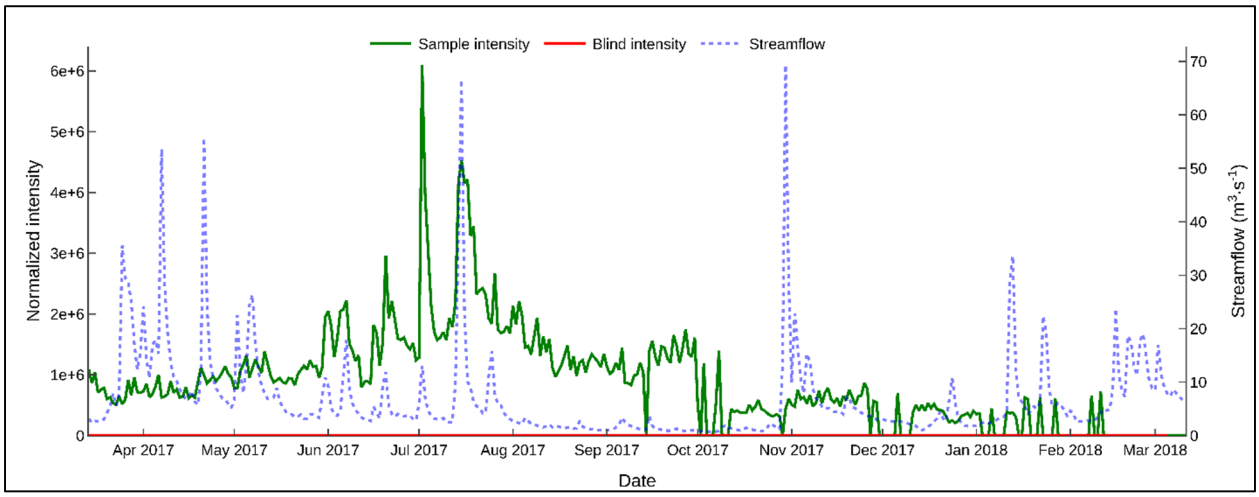



**Figure B4:** Temporal profile of atrazine-desethyl.

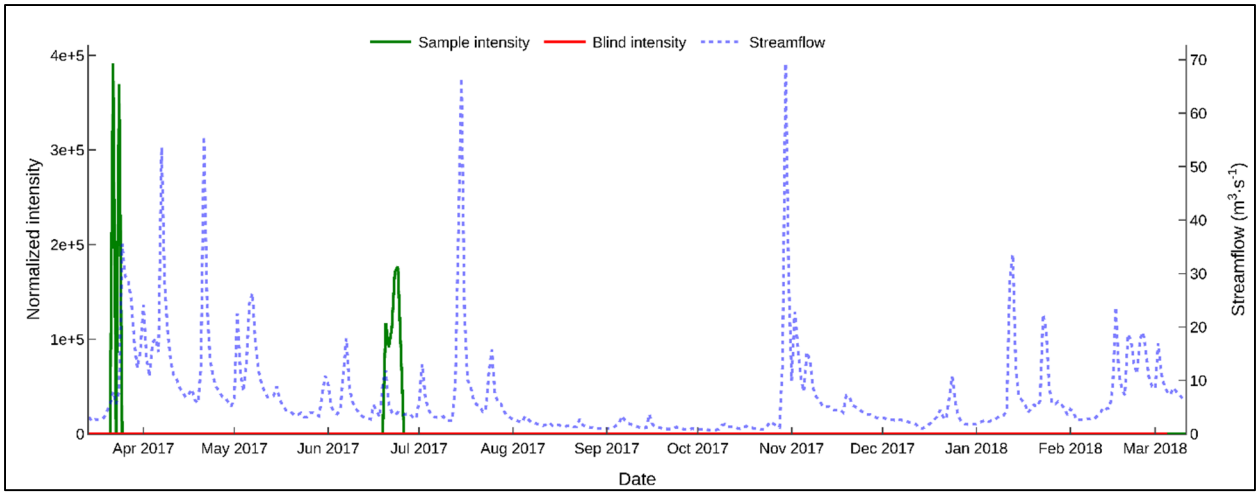



**Figure B5:** Temporal profile benzotriazole.

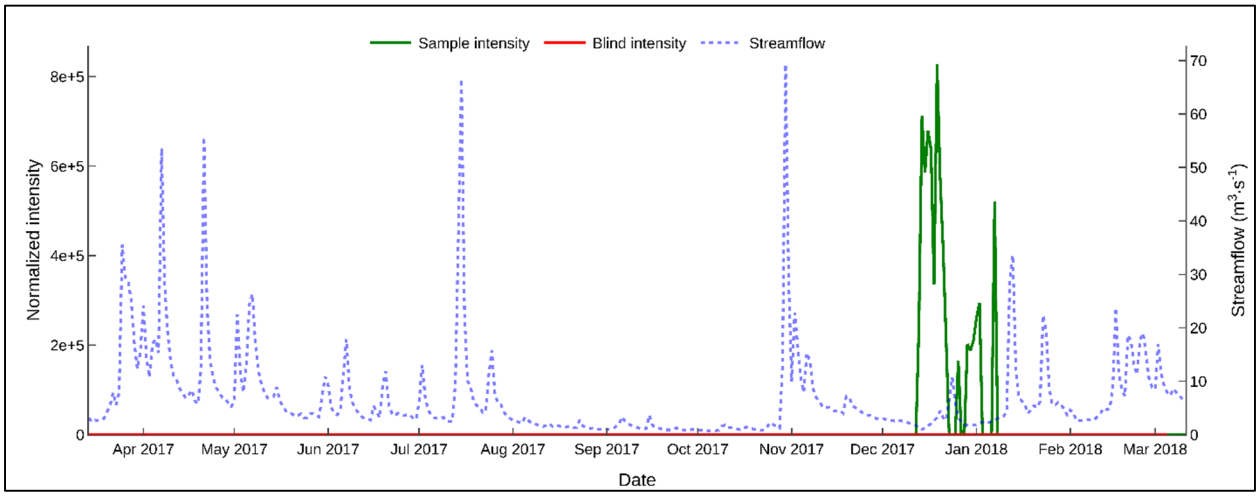



**Figure B6:** Temporal profile of bupropion.

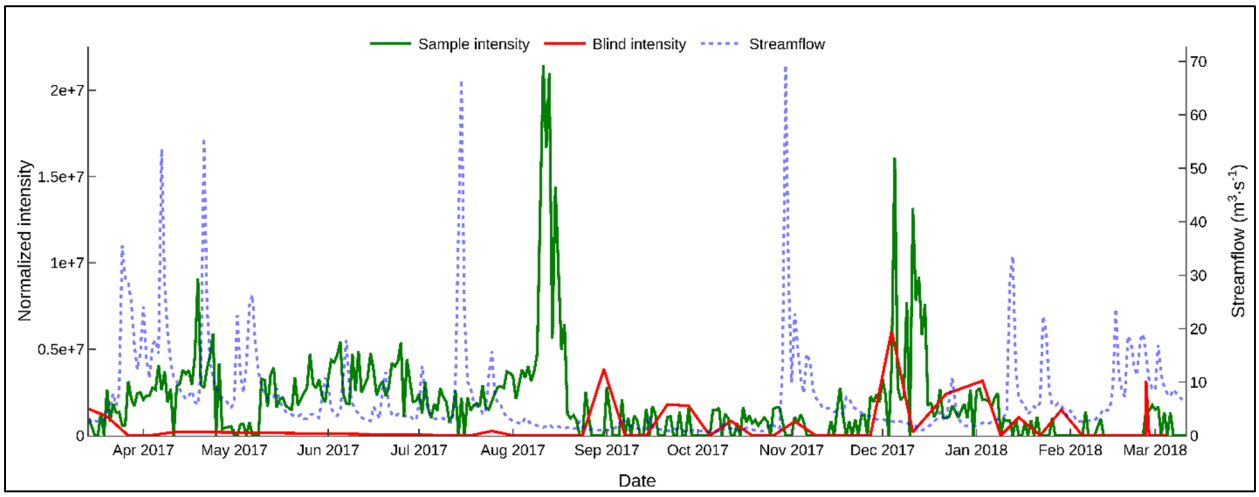



**Figure B7:** Temporal profile of caffeine.

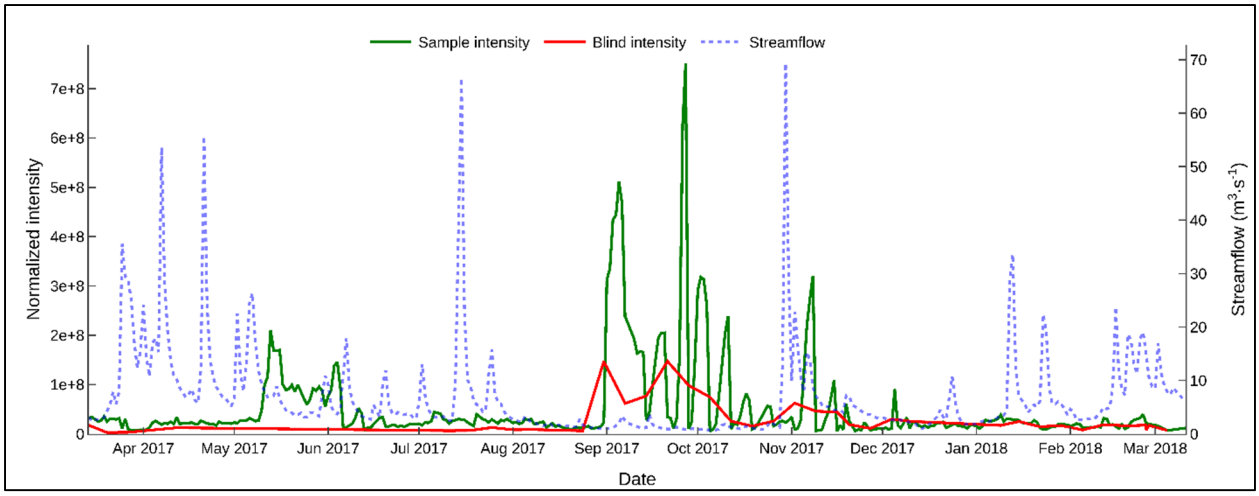



**Figure B8:** Temporal profile of diethyl phthalate.

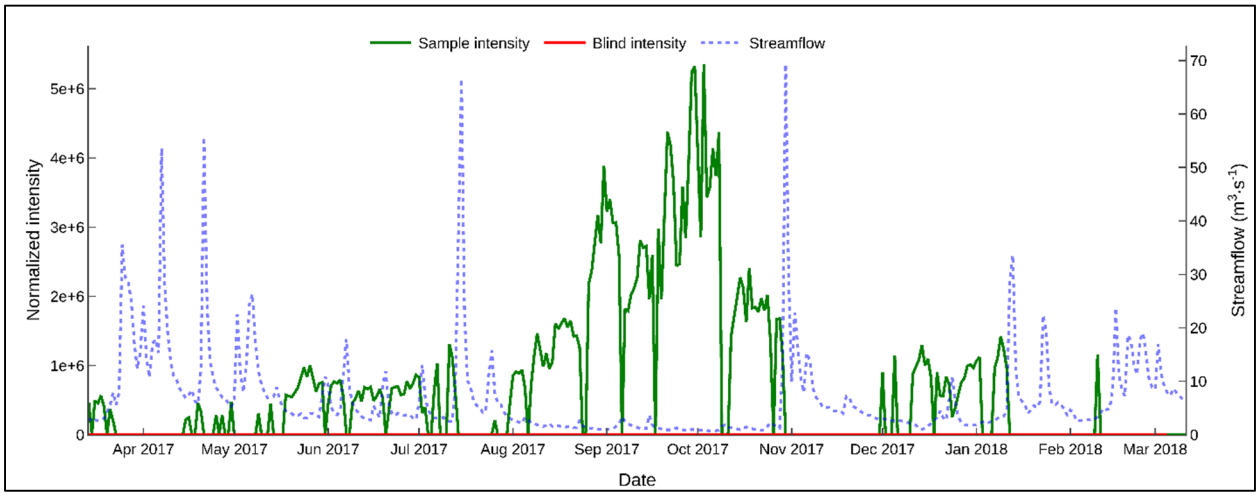



**Figure B9:** Temporal profile of fexofenadine.

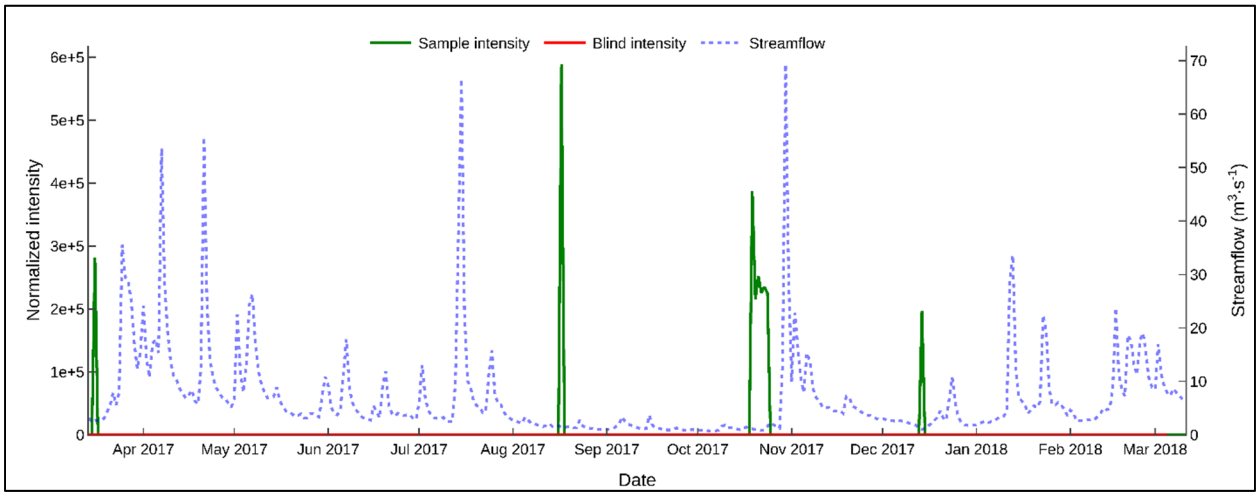



**Figure B10:** Temporal profile of fluconazole.

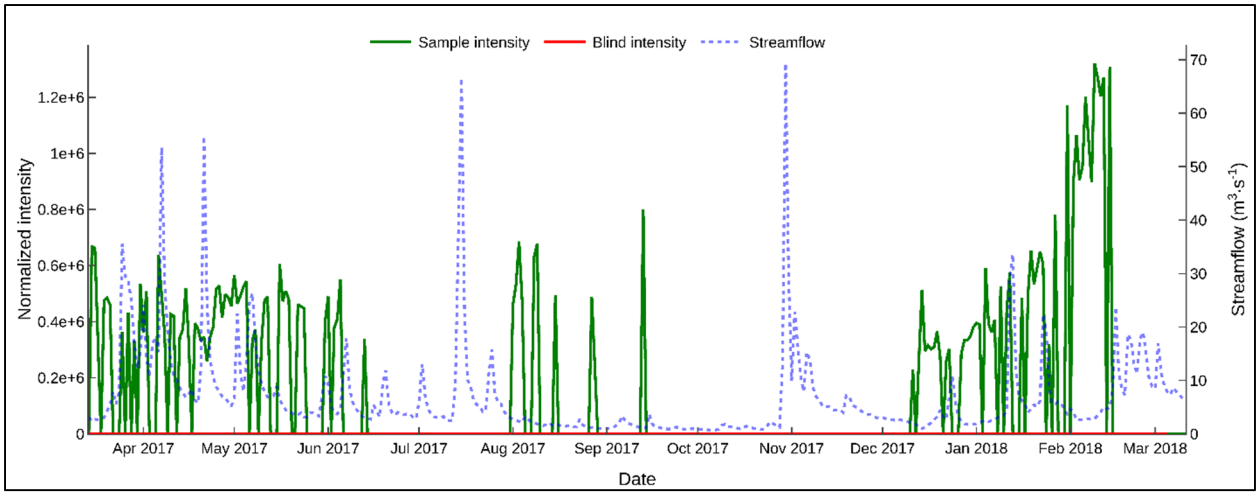



**Figure B11:** Temporal profile of gabapentin.

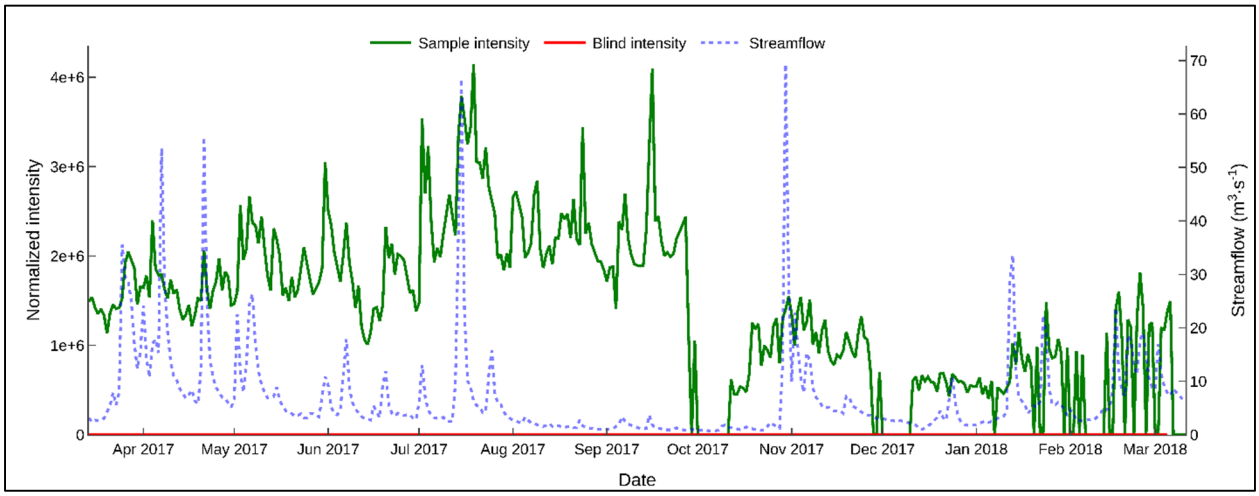



**Figure B12:** Temporal profile of hydoxy-atrazine.

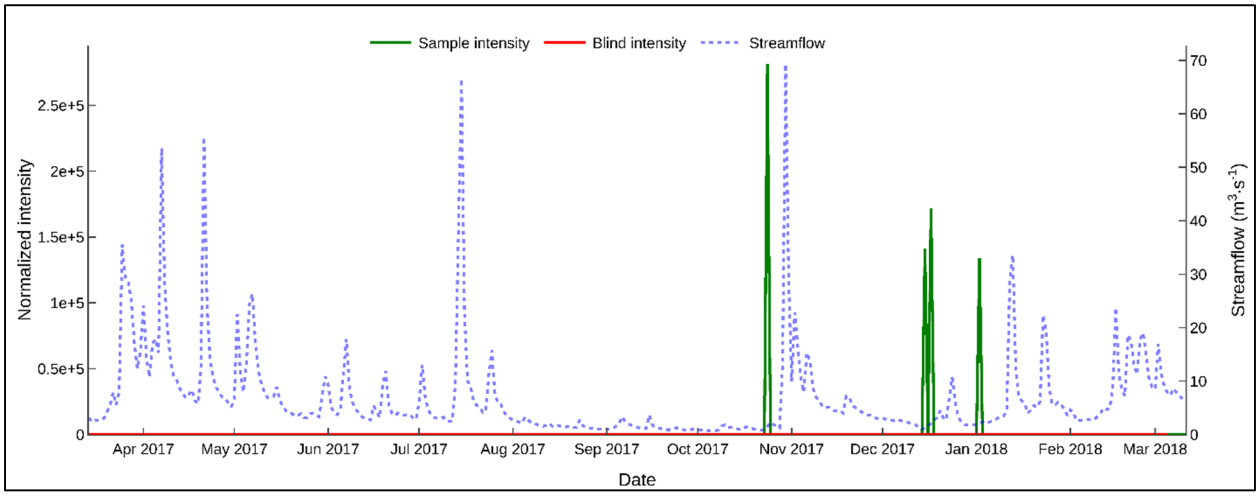



**Figure B13:** Temporal profile of irbesartan.

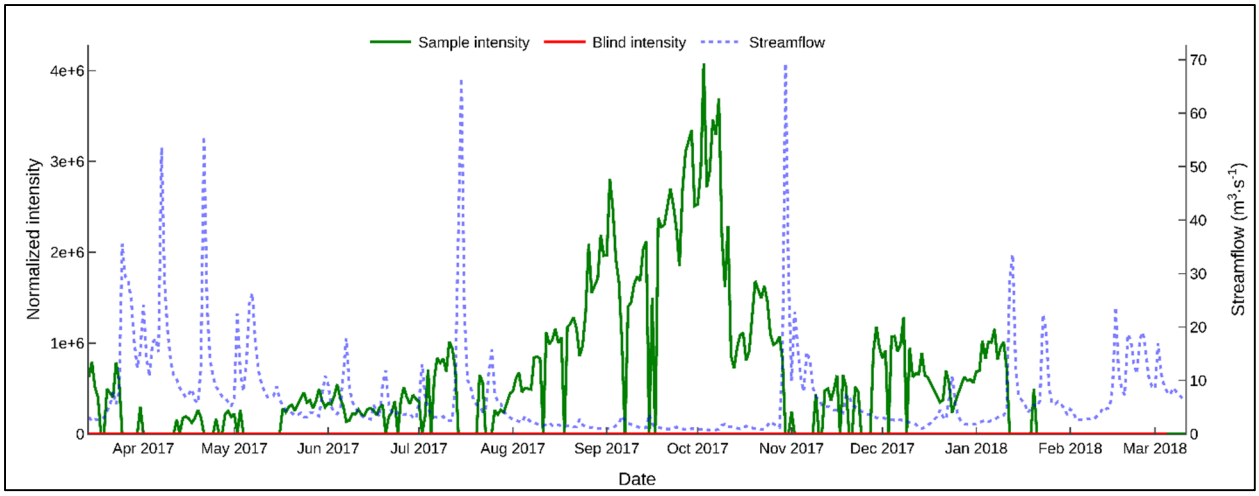



**Figure B14:** Temporal profile of lamotrigine.

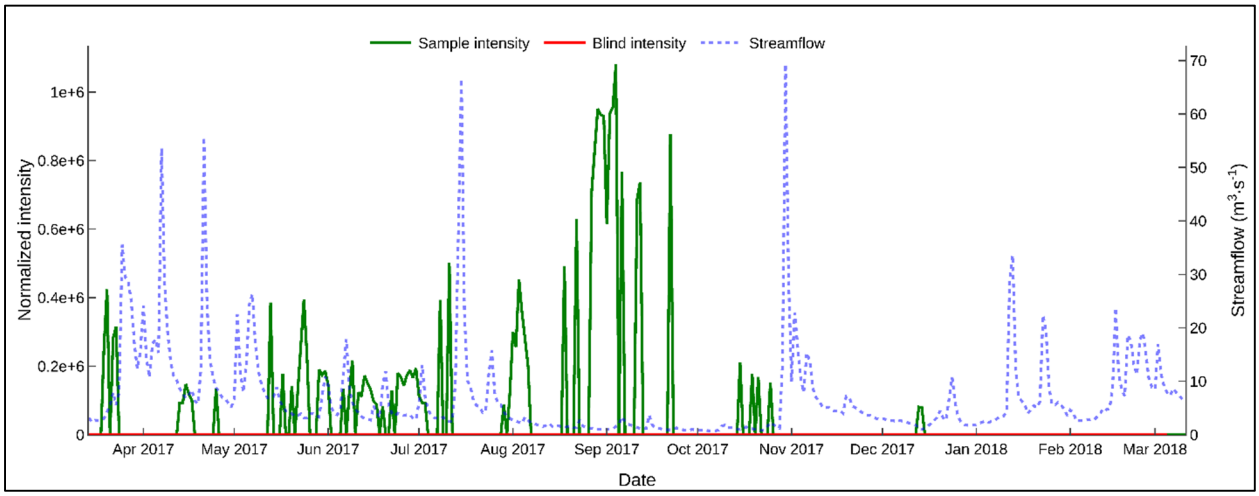



**Figure B15:** Temporal profile of lidocaine.

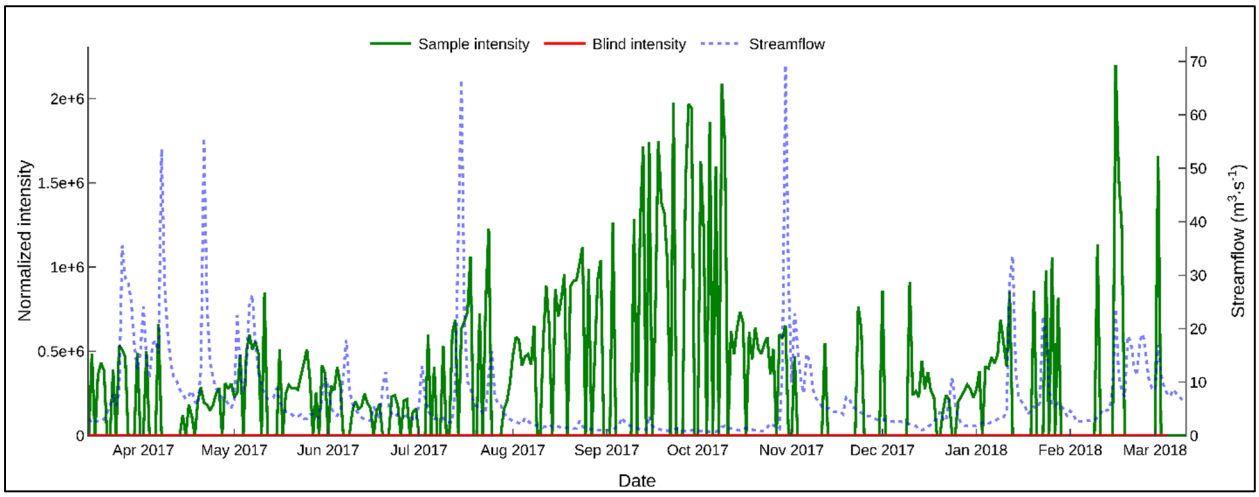



**Figure B16:** Temporal profile of losartan.

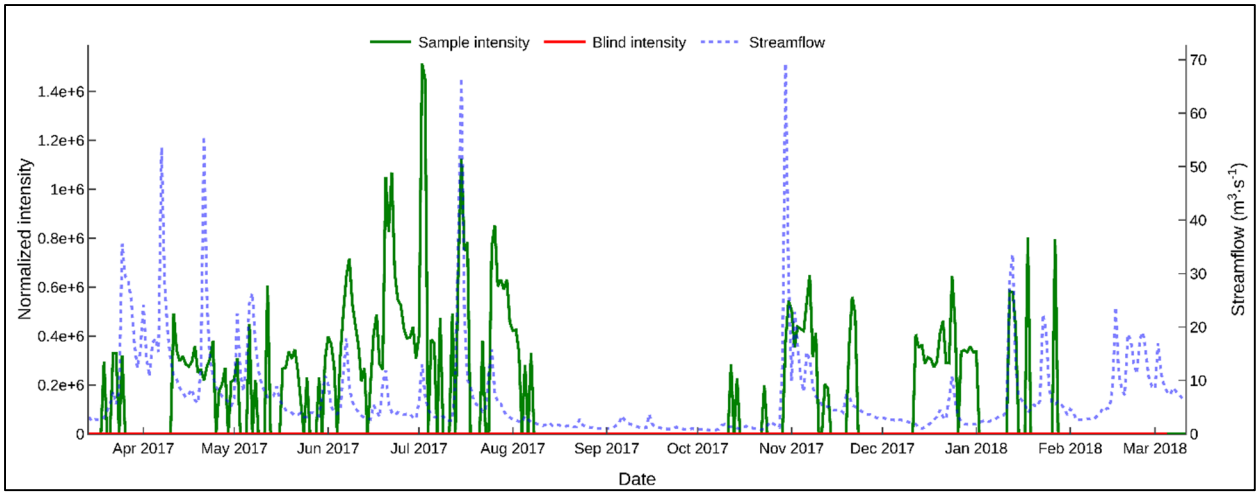



**Figure B17:** Temporal profile of metalaxyl.

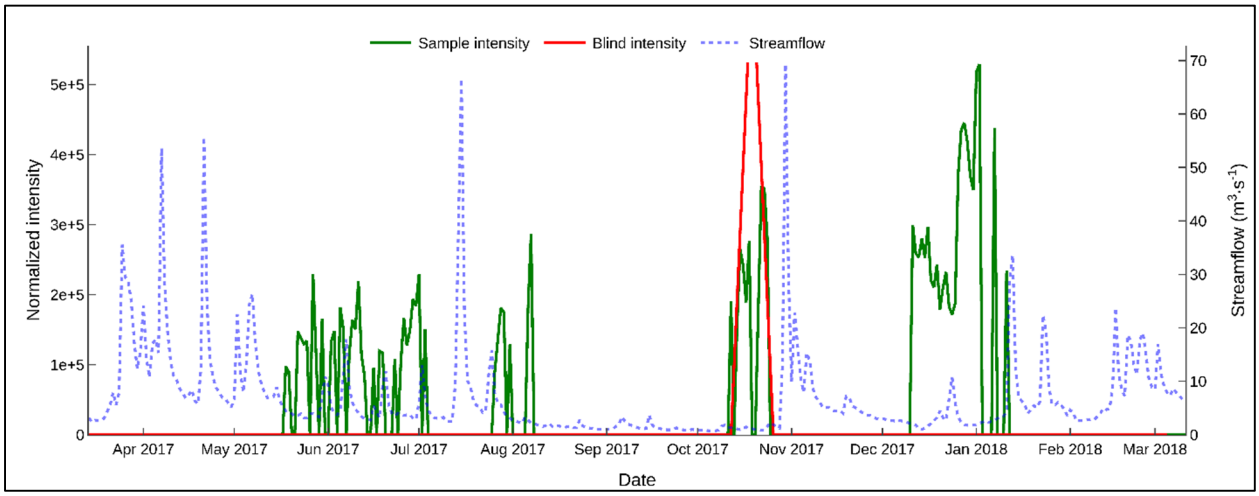



**Figure B18:** Temporal profile of methocarbamol.

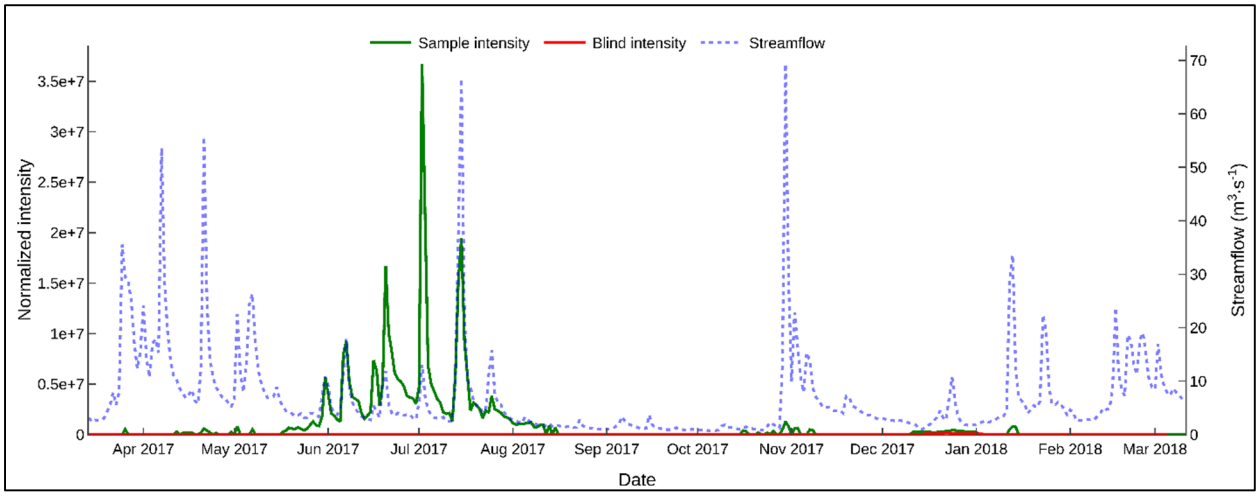



**Figure B19:** Temporal profile of metolachlor.

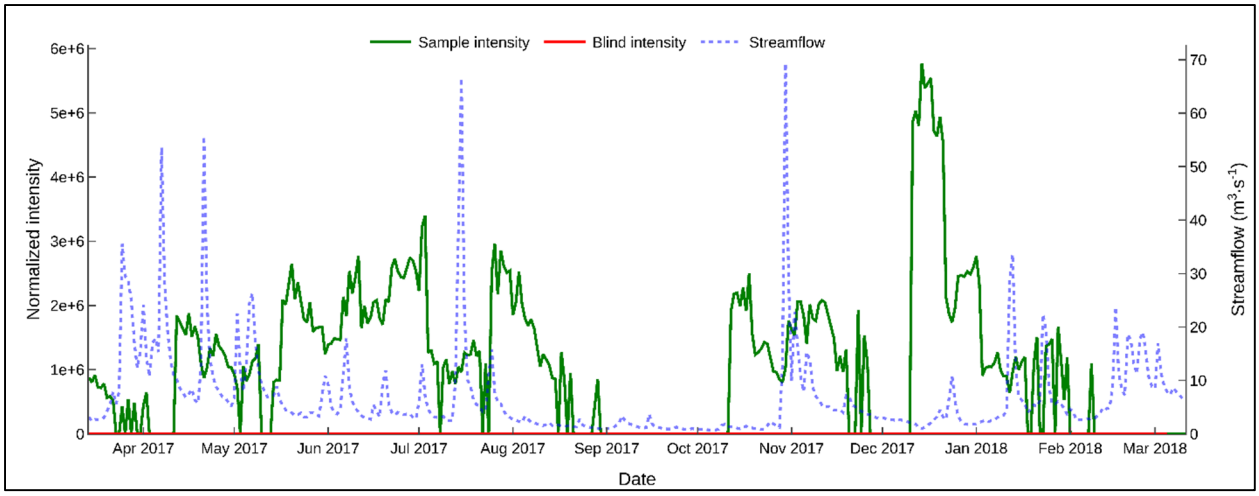



**Figure B20:** Temporal profile of metolachlor-ESA.

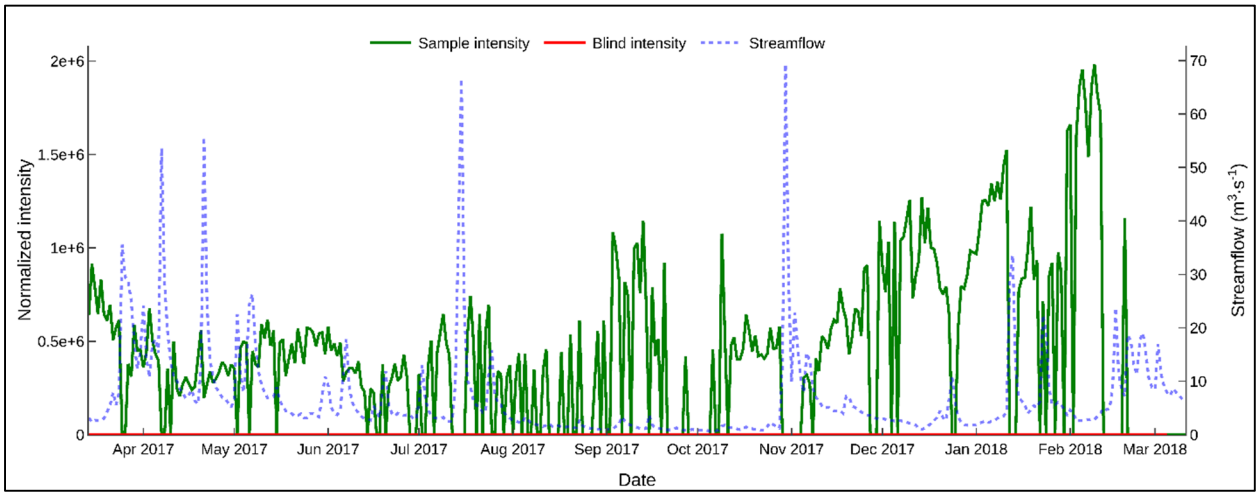



**Figure B21:** Temporal profile of metoprolol.

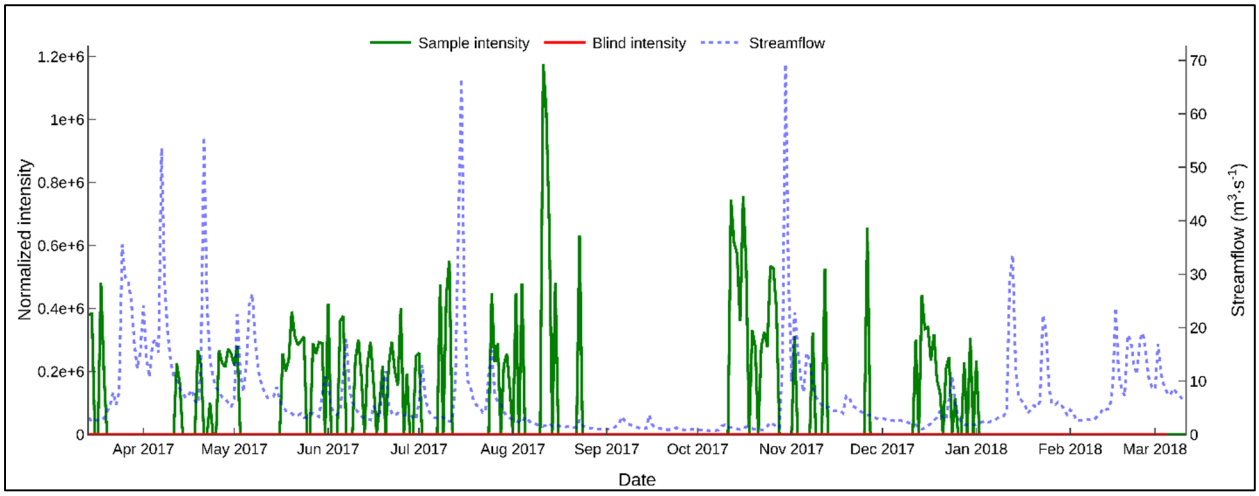



**Figure B22:** Temporal profile of prometon.

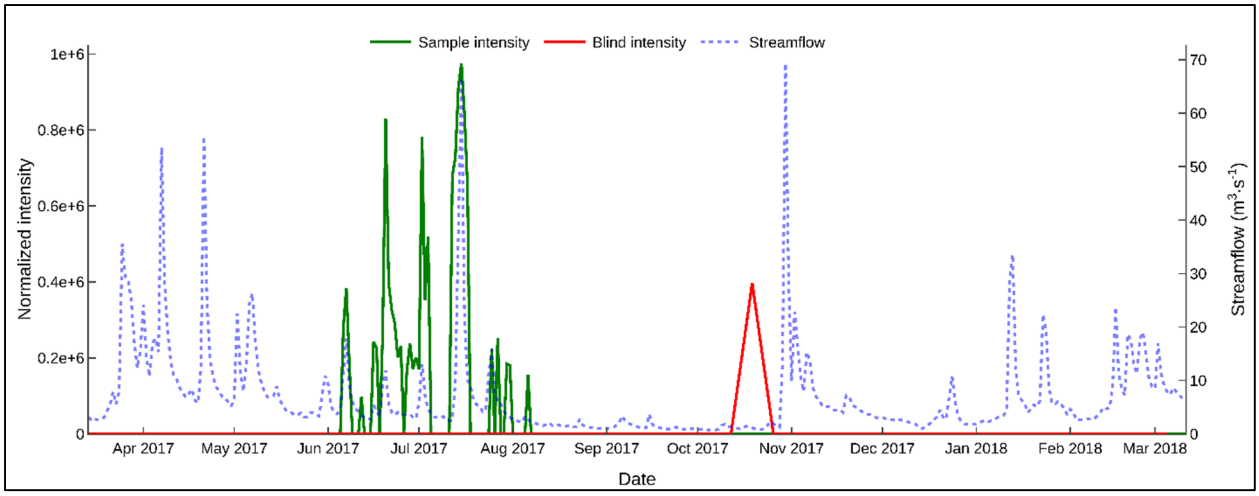



**Figure B23:** Temporal profile of propazine.

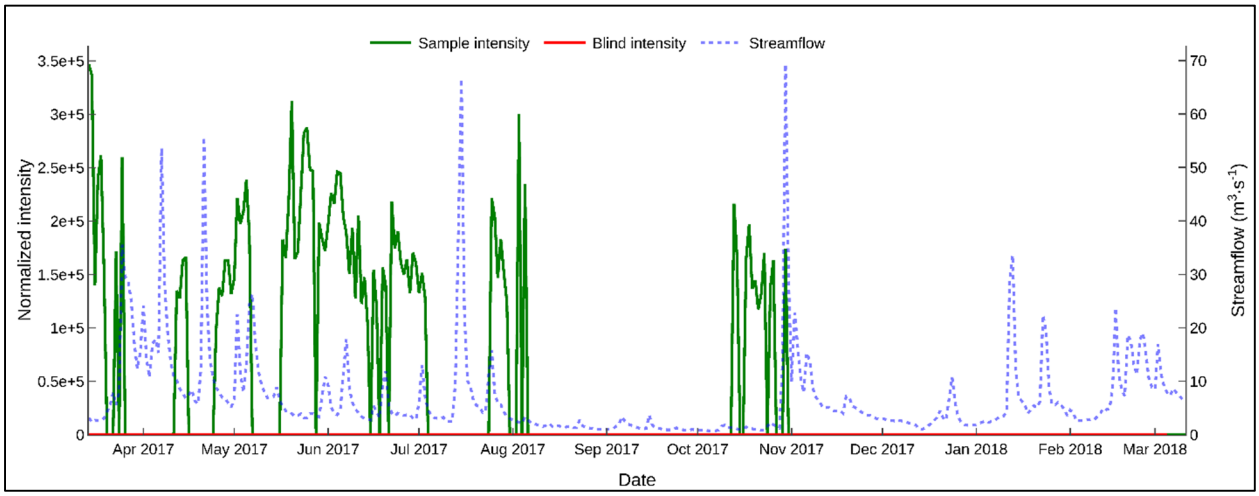



**Figure B24:** Temporal profile of ritalinic acid.

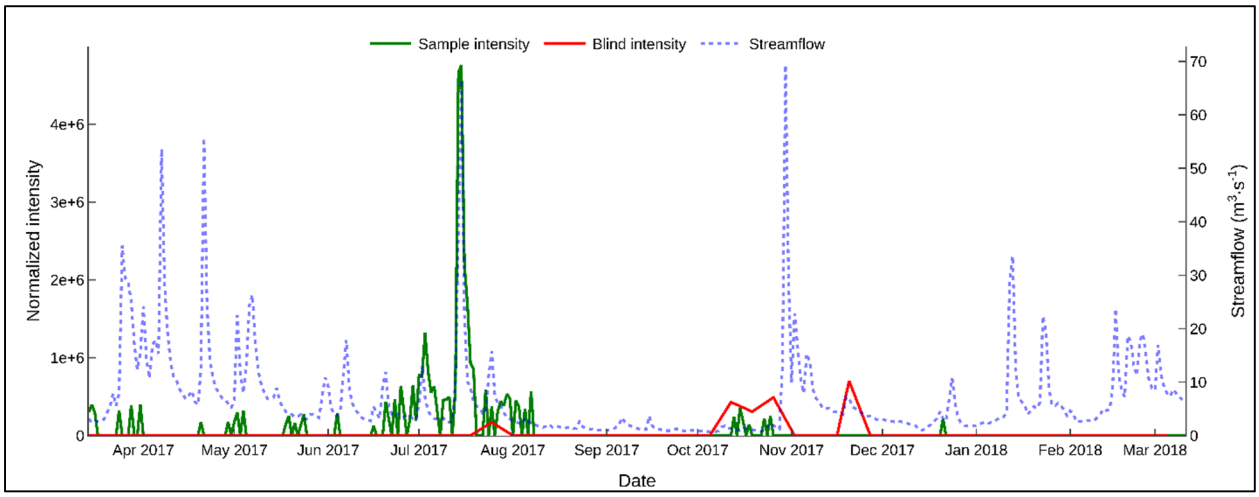



**Figure B25:** Temporal profile of simazine.

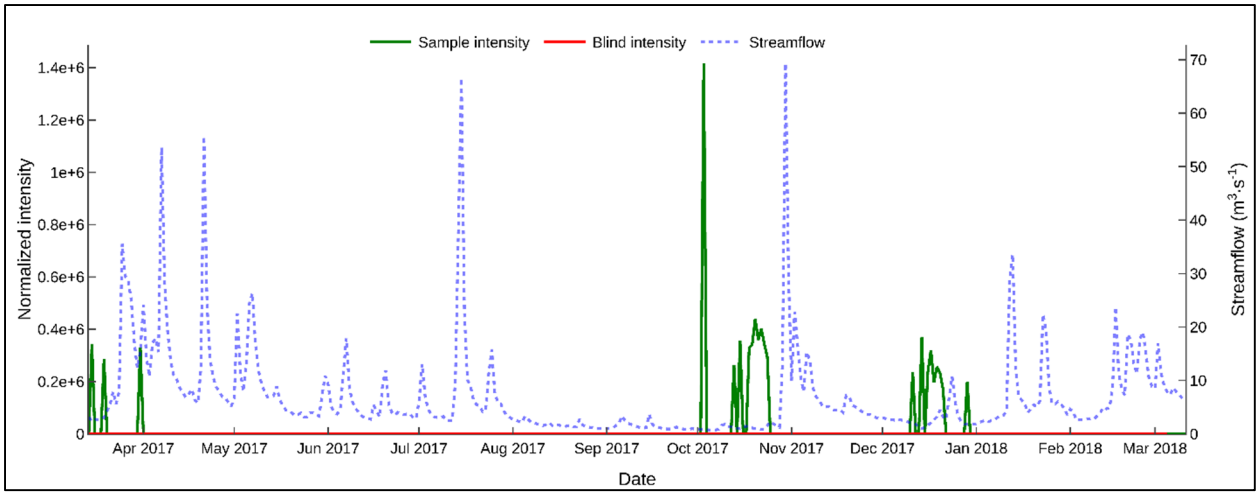



**Figure B26:** Temporal profile of sitagliptin.

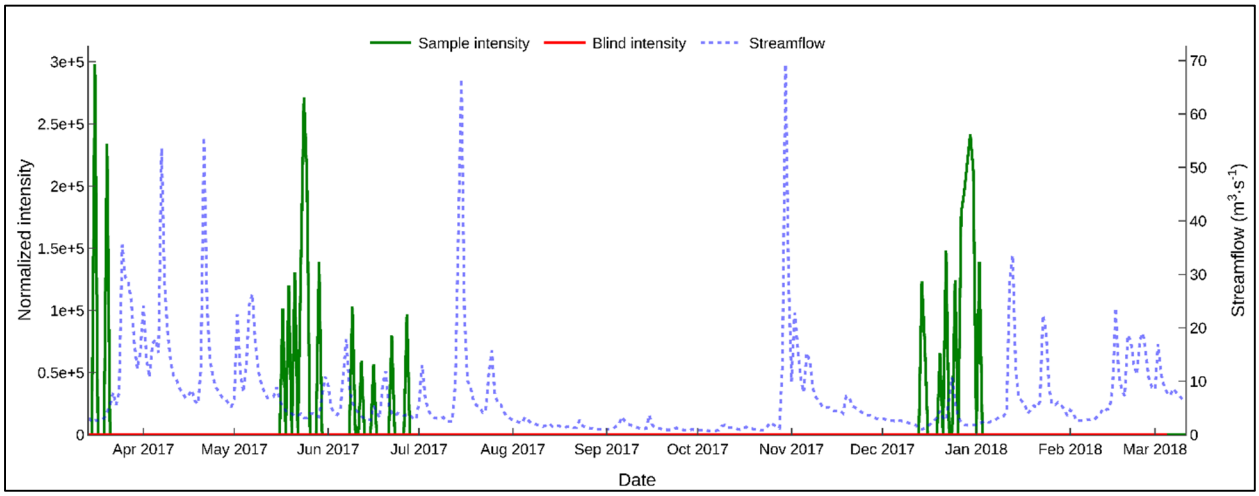



**Figure B27:** Temporal profile of trimethoprim.

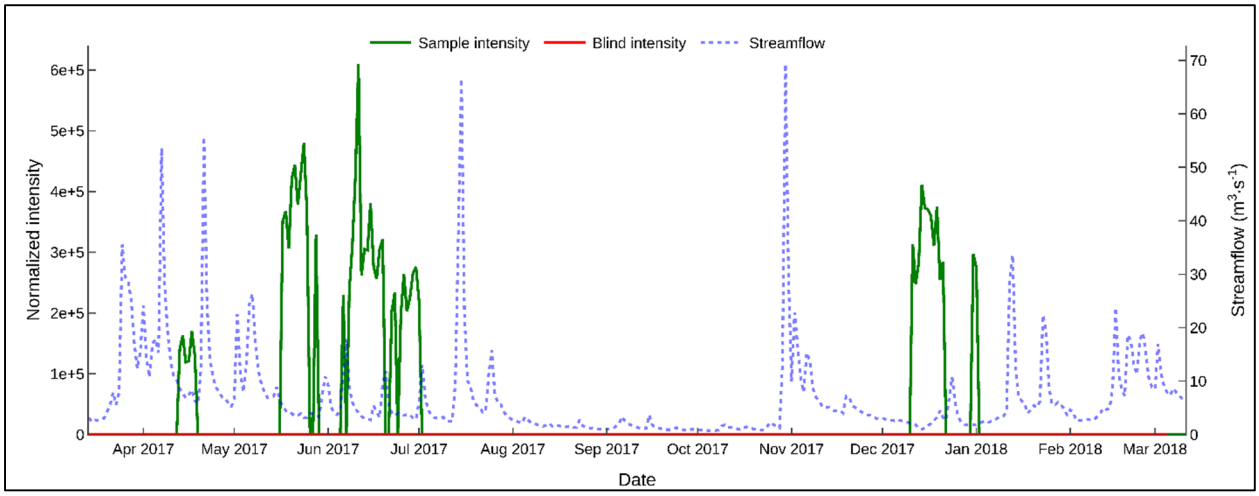



**Figure B28:** Temporal profile of valsartan.

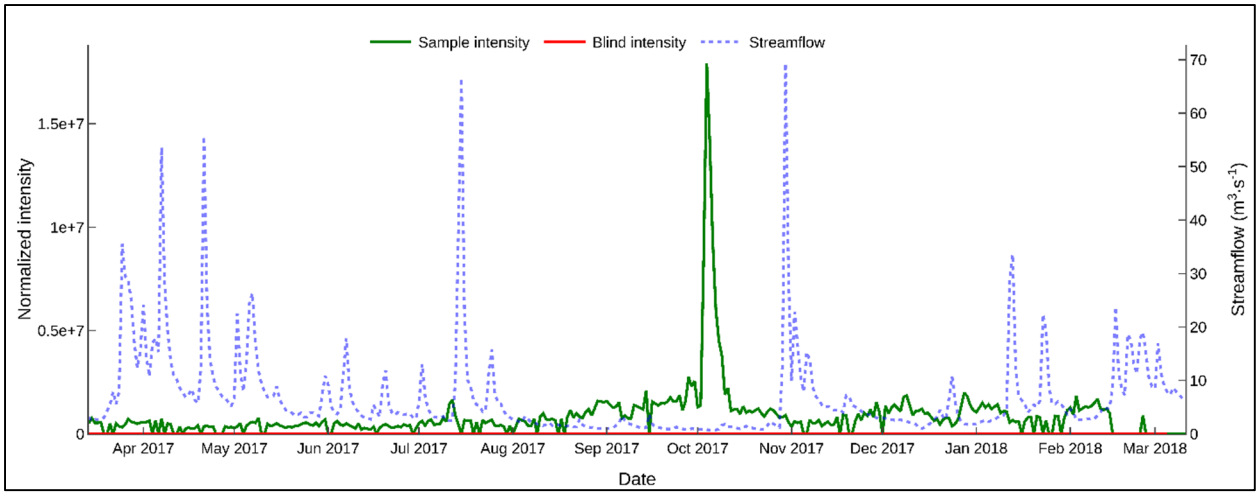



**Figure B29:** Temporal profile of venlafaxine.

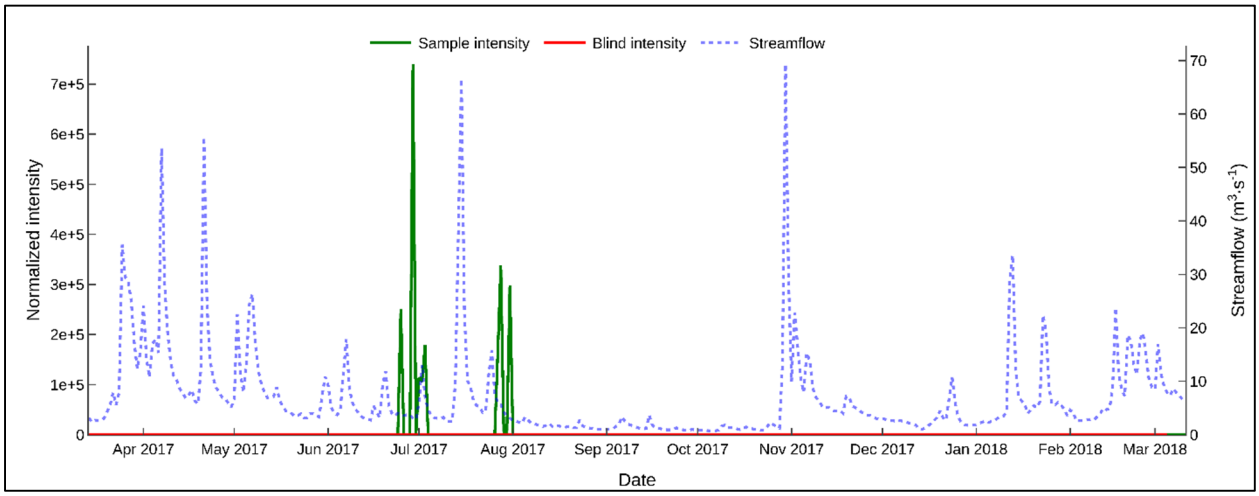



**Figure B30:** Temporal profile of warfarin.

**B.5 Filtering and clustering of MS feature profiles**

We applied a series of data reduction filters to exclude profiles that do not meet certain quality control metrics, while retaining as many target micropollutants in the final profile list as possible. The sample-to-blind ratio filter excludes profiles that were present in both the samples and field blanks. Each daily composite sample was compared with an associated weekly field blank taken throughout the study period. Through this analysis, we determined that the median sample-to-blind intensity ratio should be greater than 10 to remove background profiles caused by instrument and sample collection contamination. The mean trend intensity filter excludes profiles with low mean intensities across samples; samples with non-detects are omitted from the calculation of mean trend intensity. We found that profiles with mean trend intensities less than $10^5$ had poor isotopic signatures and noisy fragmentation patterns; all of the target compounds had mean trend intensities greater than $10^5$. Therefore, we set the mean trend intensity threshold to $10^5$. The RT filter excludes profiles with average RTs less than 6.5 min. Our results showed that these profiles mostly contained MS features with relatively poor peak shapes. We also included a data reduction step to remove profiles associated with lower order isotopologues and adducts. This step only keeps profiles associated with the main isotopologue (*e.g.,* $^{12}C$, $^{32}S$, $^{34}Cl$) and main adduct, (*e.g.*, $[M+H]^+$). This step also removes profiles that do not have any associated isotopologue or adduct profiles. This step ensures that the final profile list contains unique profiles, each associated with a unique compound or in-source fragment. Isotopologue, adduct, and in-source fragment profiles clustered very closely with the parent profiles; therefore, this step was essential to reduce the size of the final profile list and likely does not remove any relevant profiles. Lastly, a detection filter removes profiles that were detected in fewer than 30 total samples and in fewer than 10 consecutive detections. We determined that these limits considerably reduce the size of the final profile list while retaining most of the target compounds with continuous trends.

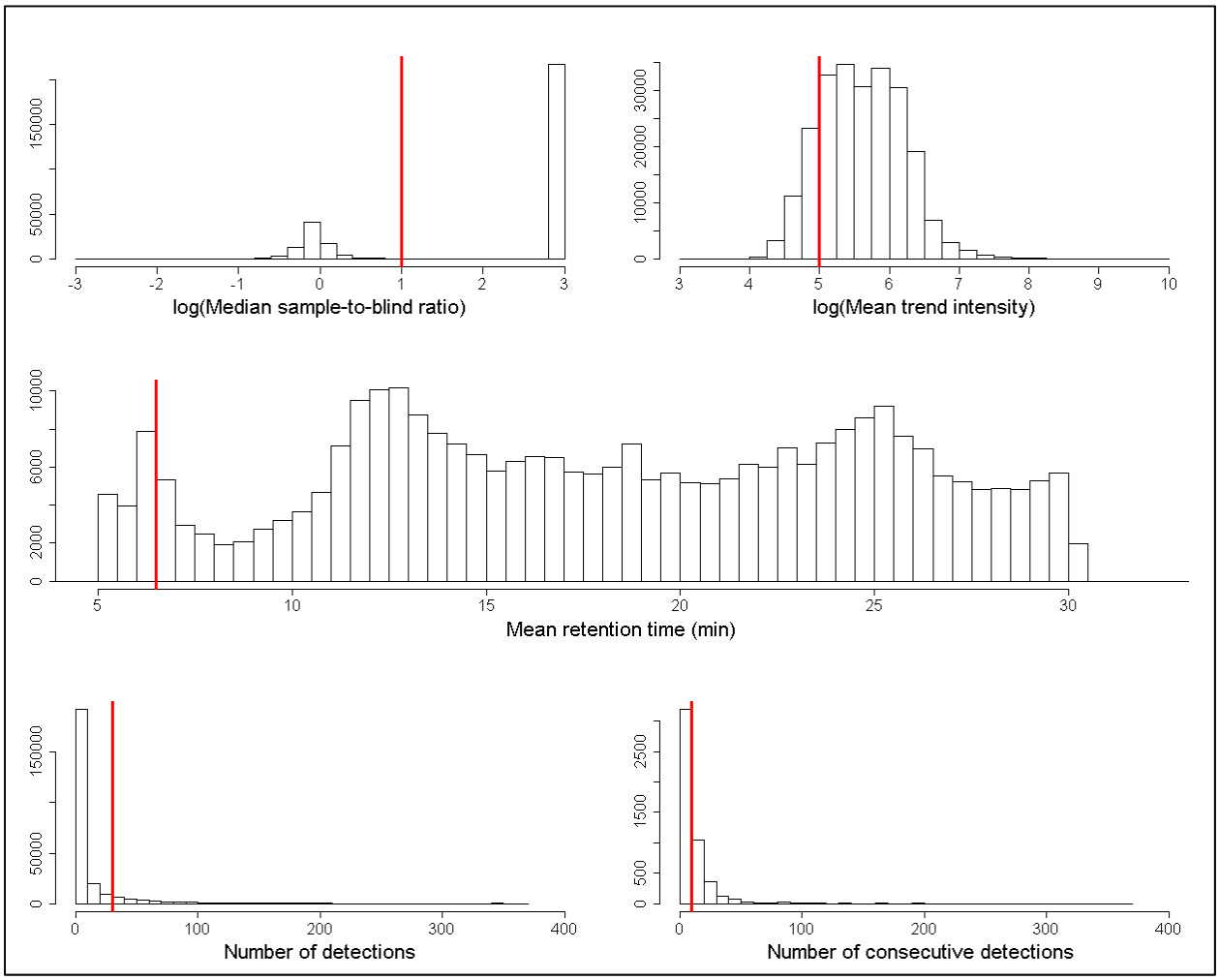



**Figure B31:** Profile frequency histograms for each data reduction filter. Red lines indicate cutoff values for each filter: median sample-to-blind ratio ≥ 10 (profiles without detections in the blinds were assigned a value of $10^3$), mean trend intensity ≥ $10^5$, mean retention time ≥ 6.5 min, number of detections ≥ 30, and number of consecutive detections ≥ 10.

## B.7 Structural elucidation of nontarget MS features

The weighting factors for the candidate structure ranking were iteratively optimized to maximize the number of true positive top-ranked candidates of the 31 target micropollutants previously identified. An in-depth discussion of the weighting factors is provided elsewhere and the weighting factors used in our study were initially set to match previously optimized literature values (fragment score = 0.50, RT score = 0.16, and reference (PubChem) score = 0.34),[10] which resulted in a true positive rate of 81% (25 of 31). To improve the accuracy of the workflow, we added additional weighting factors and changed their relative weights. We split the fragment score into two separate fragment-related scores: the fragment score (how well the measured $MS^2$ fragments are explained by the *in silico* $MS^2$ fragments) and the Metfusion score (how well the measured $MS^2$ fragments match with fragments of similar structures in the MoNA spectral library). We also found that by including a suspect score, the rankings improved by providing higher scores to more relevant candidates. Our final weighting factors (fragment score = 0.30, Metfusion score = 0.30, PubChem score = 0.10 (references = 0.05 and patents = 0.05), RT score = 0.15, and suspect score = 0.15) resulted in a true positive rate of 97% (30 of 31). The sole target micropollutant that was not the top ranked candidate was desvenlafaxine, which was ranked as the #2 candidate behind tramadol, a structural isomer.

Temporal profiles for the 59 nontarget micropollutants that were identified in this study are provided in **Figures B32 – S90** and are organized in the order in which they were clustered (from left to right) in the dendrogram provided in **Figure 3.2**.

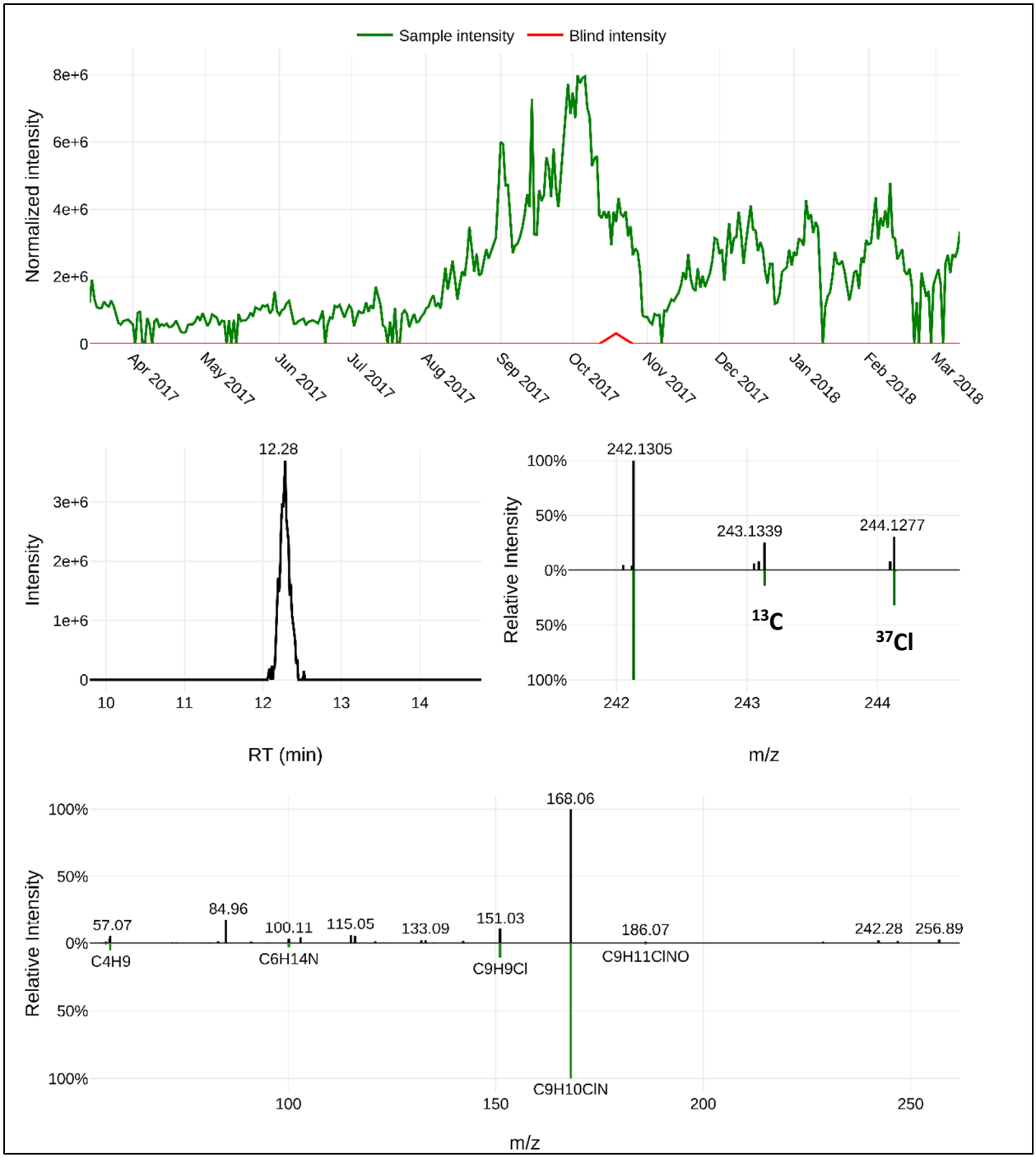



**Figure B32:** Identification of rac-threo-dihydrobupropion – level 1.

Rac-threo-dihydrobupropion ($C_{13}H_{20}ClNO$) was confirmed using an authentic reference standard which matched the $RT_{avg}$ of 12.2 min, the MS spectra ($m/z_{avg}$ = 242.1306 for $[M+H]^+$, $\Delta m$ = 0.13 ppm), the theoretical abundance (13%) of the $^{13}C$ monoisotopic mass, the theoretical abundance (32%) of the $^{37}Cl$ monoisotopic mass, and the $MS^2$ fragments ($m/z$ = 57.07, 151.03, and 168.06).

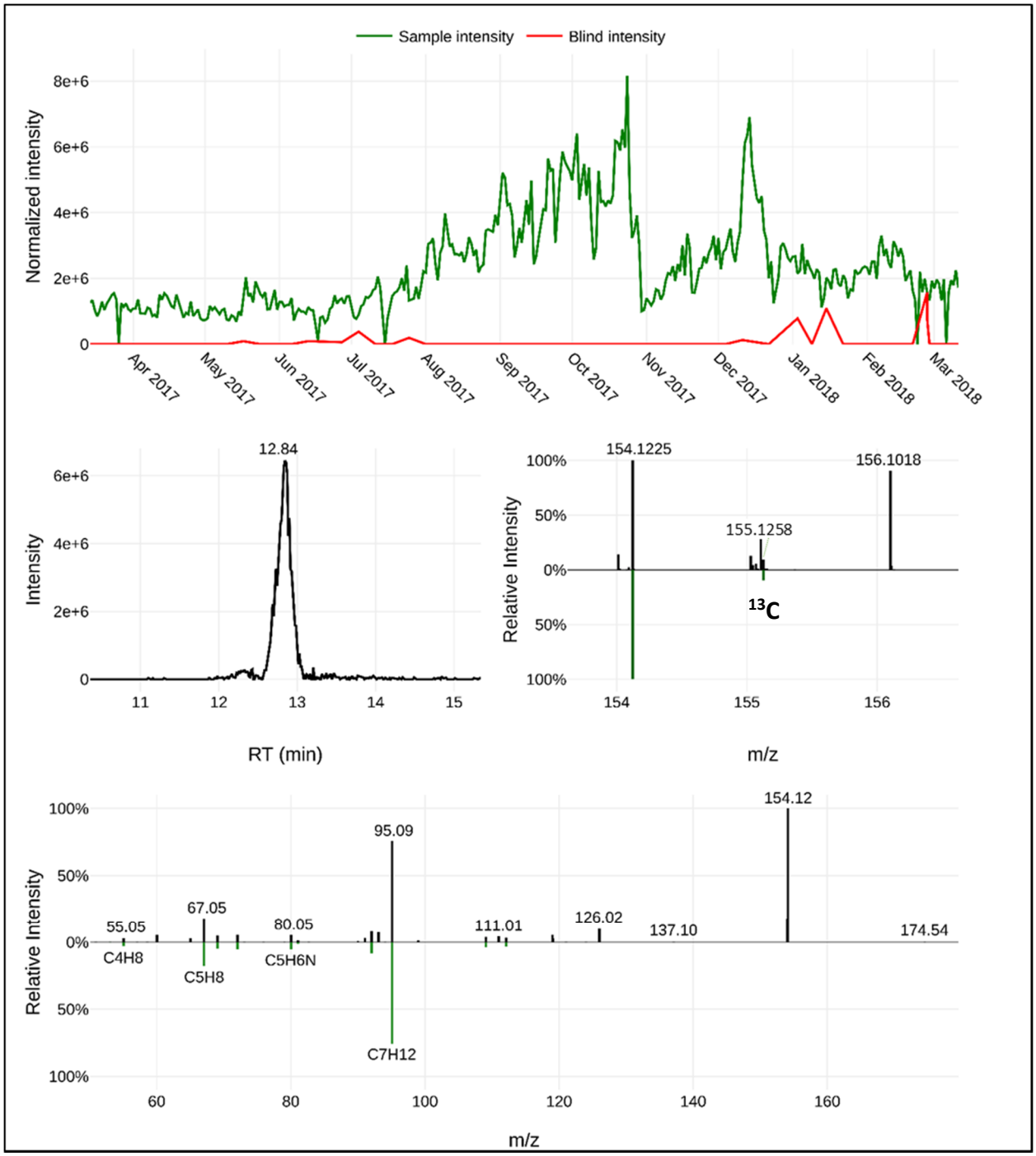



**Figure B33:** Identification of gabapentin-lactam – level 1.

Gabapentin-lactam ($C_9H_{15}NO$) was confirmed using an authentic reference standard which matched the $RT_{avg}$ of 12.8 min, the MS spectra ($m/z_{avg}$ = 154.1226 for $[M+H]^+$, Δm = -0.57 ppm), the theoretical abundance (9%) of the $^{13}C$ monoisotopic mass, and the $MS^2$ fragments ($m/z$ = 67.05, 93.07, and 95.09).

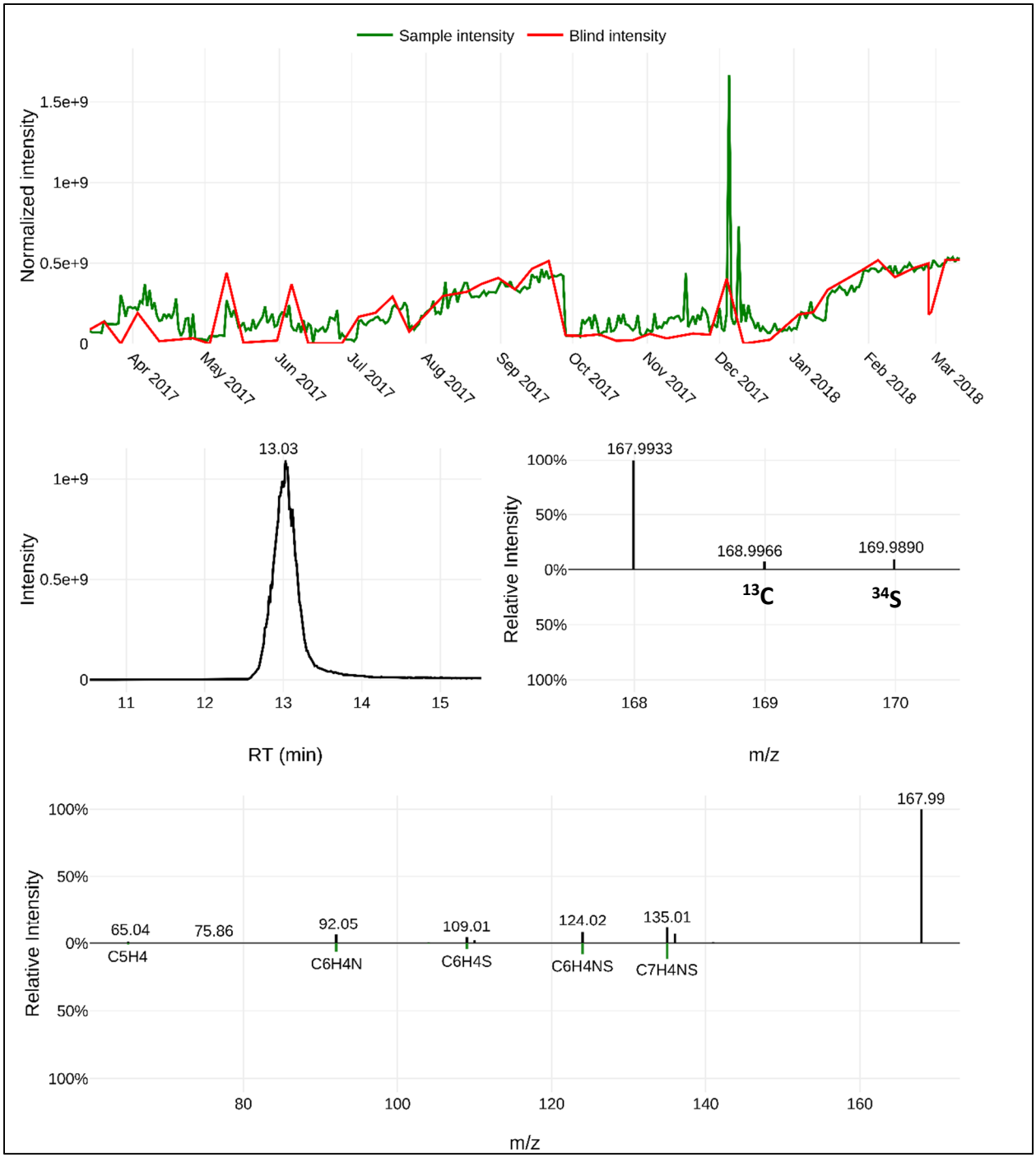



**Figure B34:** Identification of 2-mercaptobenzothiazole – level 1.

2-Mercaptobenzothiazole ($C_7H_5NS_2$) was confirmed using an authentic reference standard which matched the $RT_{avg}$ of 12.9 min, the MS spectra ($m/z_{avg}$ = 167.9936 for $[M+H]^+$, $\Delta m$ = -0.28 ppm), the theoretical abundance (7%) of the $^{13}C$ monoisotopic mass, the theoretical abundance (9%) of the $^{34}S$ monoisotopic mass, and the $MS^2$ fragments ($m/z$ = 109.01, 124.02, and 135.01).

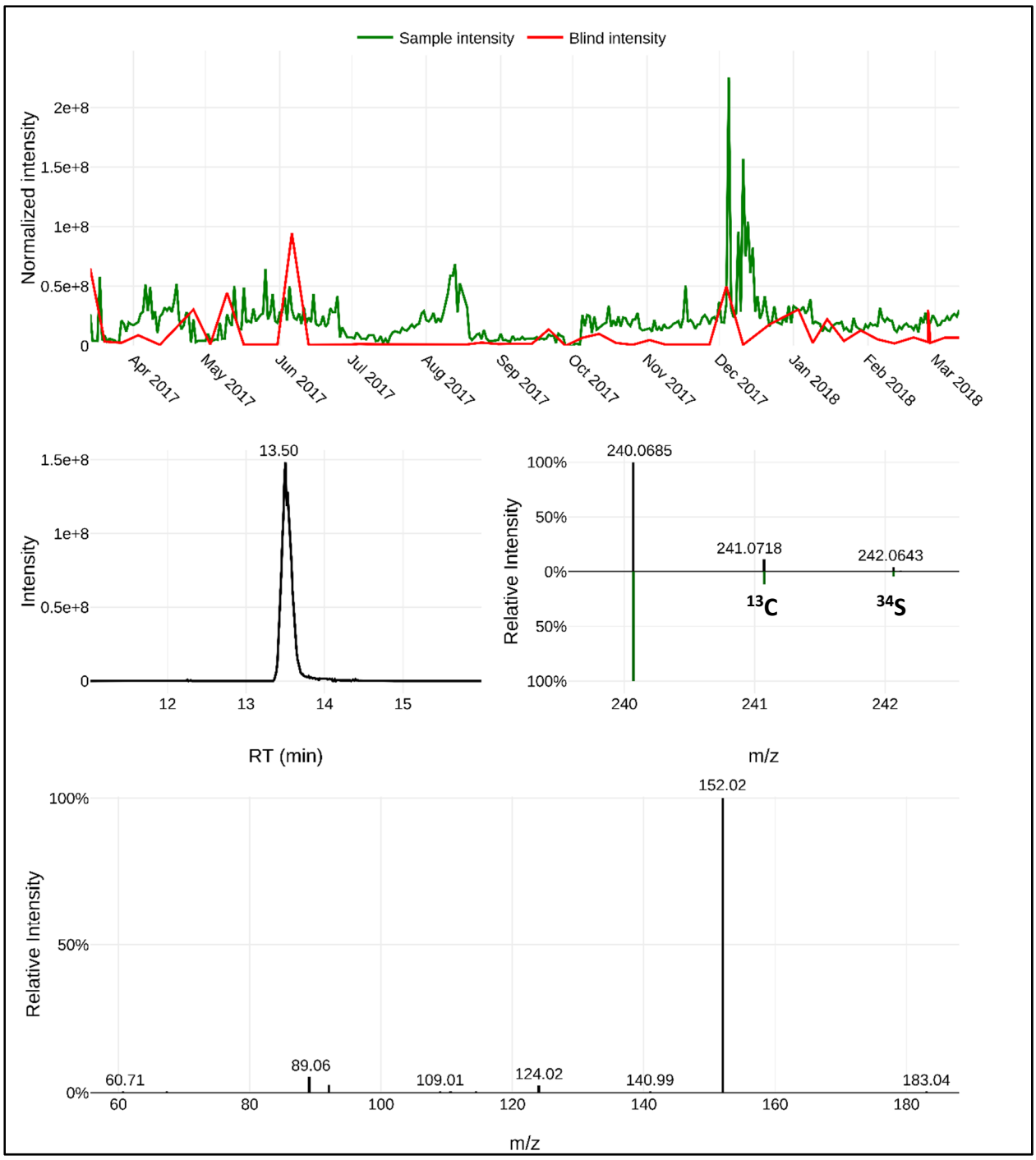



**Figure B35:** Identification of NT240 – level 3.

NT240 was identified to have an unequivocal molecular formula ($C_{11}H_{13}NO_3S$) which had a $RT_{avg}$ of 13.6 min. The proposed molecular formula assignment matches the MS spectra ($m/z_{avg}$ = 240.0689 for $[M+H]^+$, $\Delta m$ = 0.09 ppm), the theoretical abundance (11%) of the $^{13}C$ monoisotopic mass, the theoretical abundance (4.5%) of the $^{34}S$ monoisotopic mass, and the $MS^2$ fragments. The $MS^2$ fragments ($m/z$ = 124.02, and 152.02) suggest a benzothiazolone sub-structure.

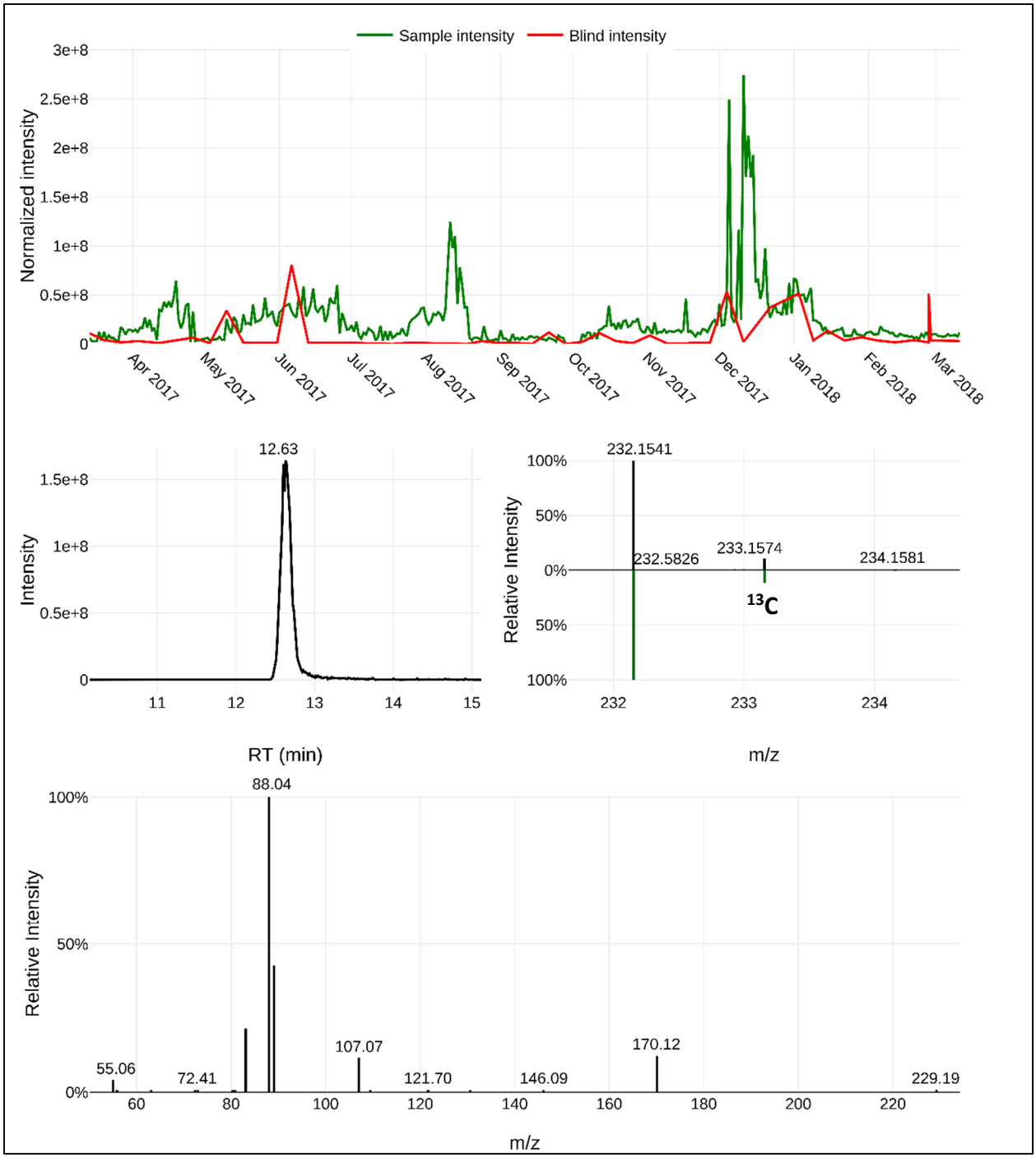



**Figure B36:** Identification of NT232 – level 4.

NT232 was identified to have an unequivocal molecular formula ($C_{11}H_{21}NO_4$) which had a $RT_{avg}$ of 12.7 min. The proposed molecular formula assignment matches the MS spectra ($m/z_{avg}$ = 232.1544 for $[M+H]^+$, $\Delta m$ = 0.19 ppm), the theoretical abundance (11%) of the $^{13}C$ monoisotopic mass, and the $MS^2$ fragments.

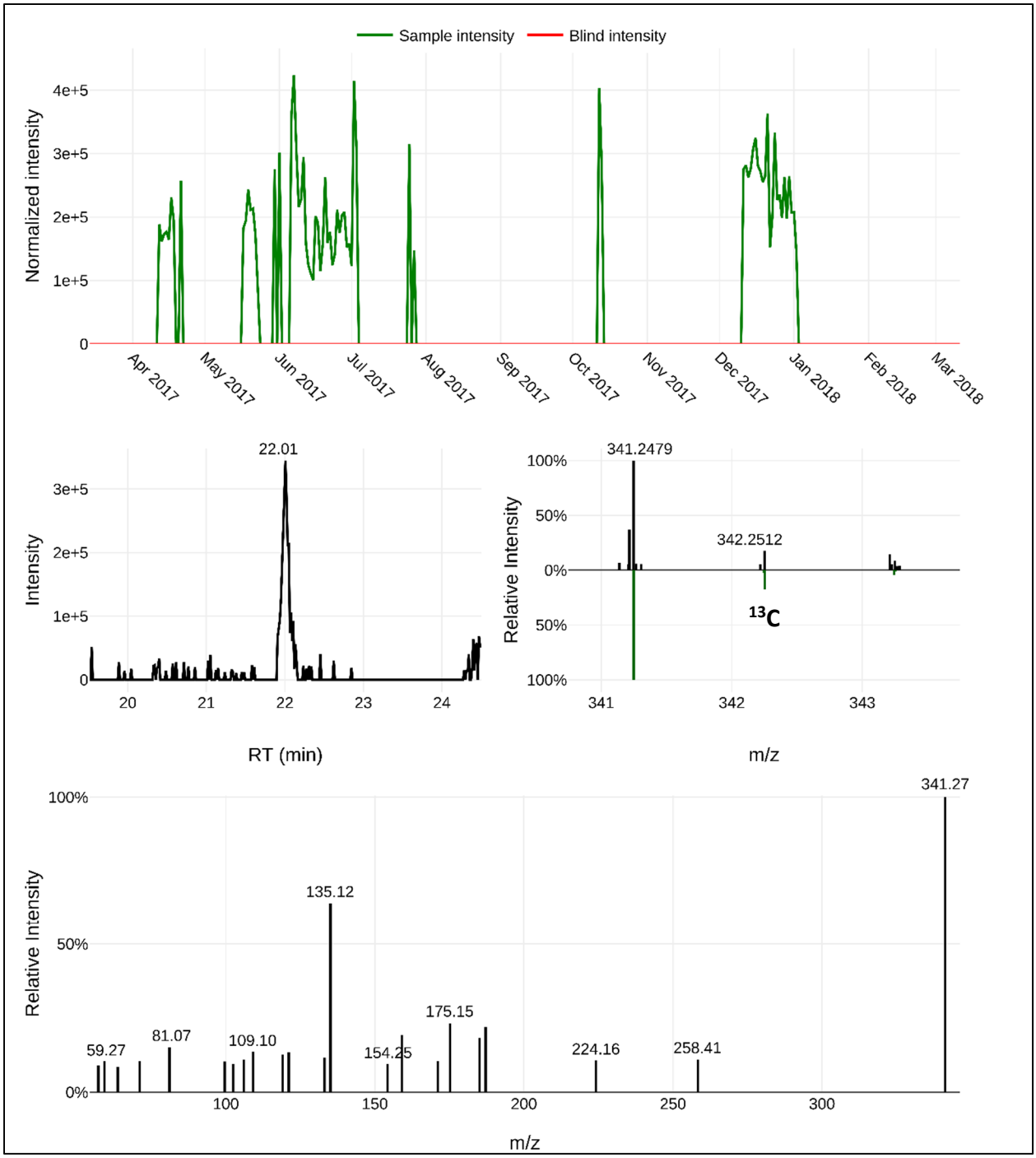



**Figure B37:** Identification of NT341 – level 4.

NT341 was identified to have an unequivocal molecular formula ($C_{23}H_{32}O_2$) which had a $RT_{avg}$ of 22.1 min. The proposed molecular formula assignment matches the MS spectra ($m/z_{avg}$ = 341.2474 for $[M+H]^+$, $\Delta m$ = -0.29 ppm), the theoretical abundance (23%) of the $^{13}C$ monoisotopic mass, and the $MS^2$ fragments.

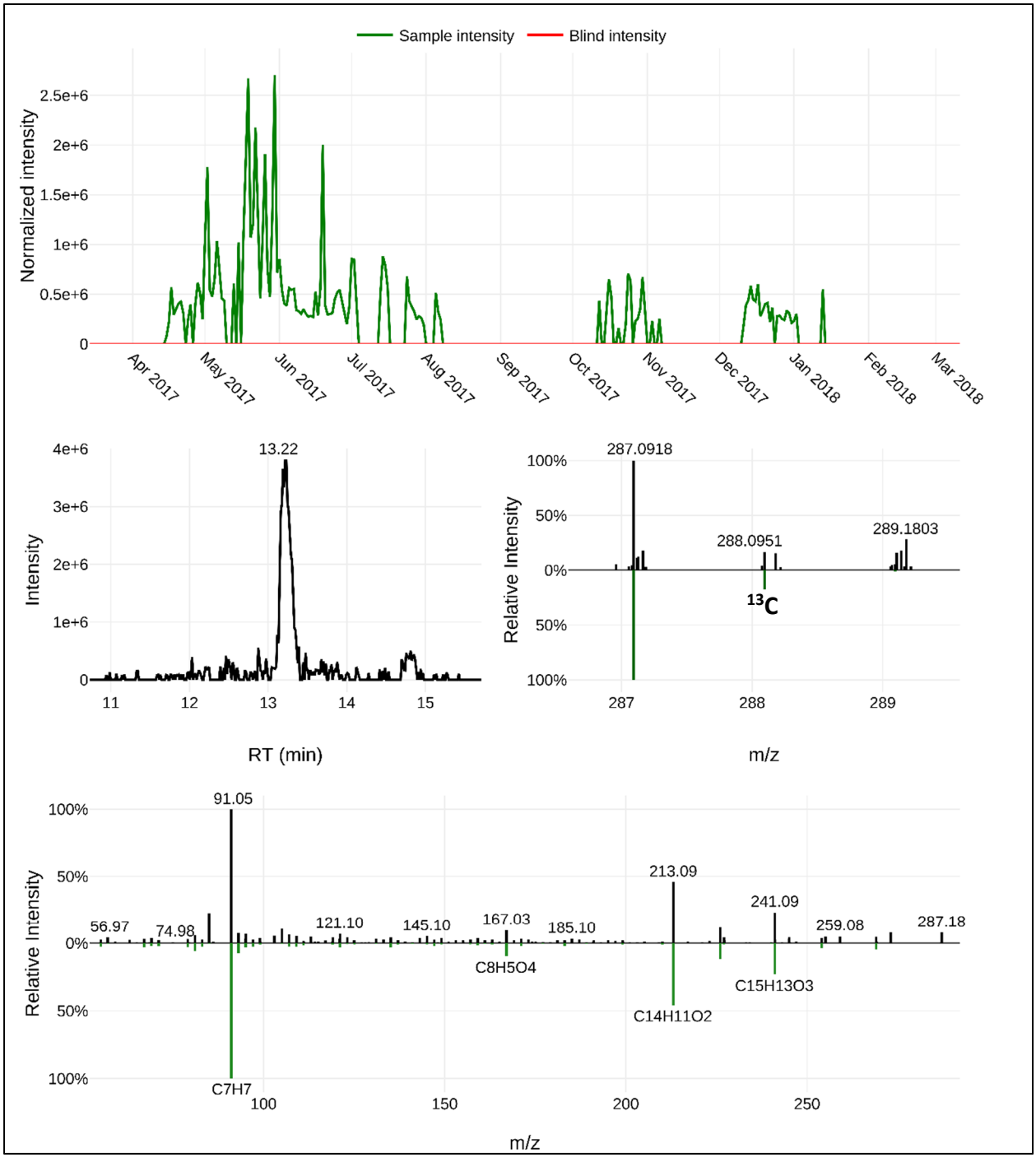



**Figure B38:** Identification of NT287 – level 2P.

NT287 likely represents 8-benzyloxy-2,3-dihydro-1,4-benzodioxin-5-carboxylic acid ($C_{16}H_{14}O_5$) which had a $RT_{avg}$ of 13.2 min. The proposed MS assignment matches the MS spectra ($m/z_{avg}$ = 287.0915 for $[M+H]^+$, $\Delta m$ = 0.31 ppm), the theoretical abundance (16%) of the $^{13}C$ monoisotopic mass, and the *in silico* $MS^2$ fragments generated with Metfrag.

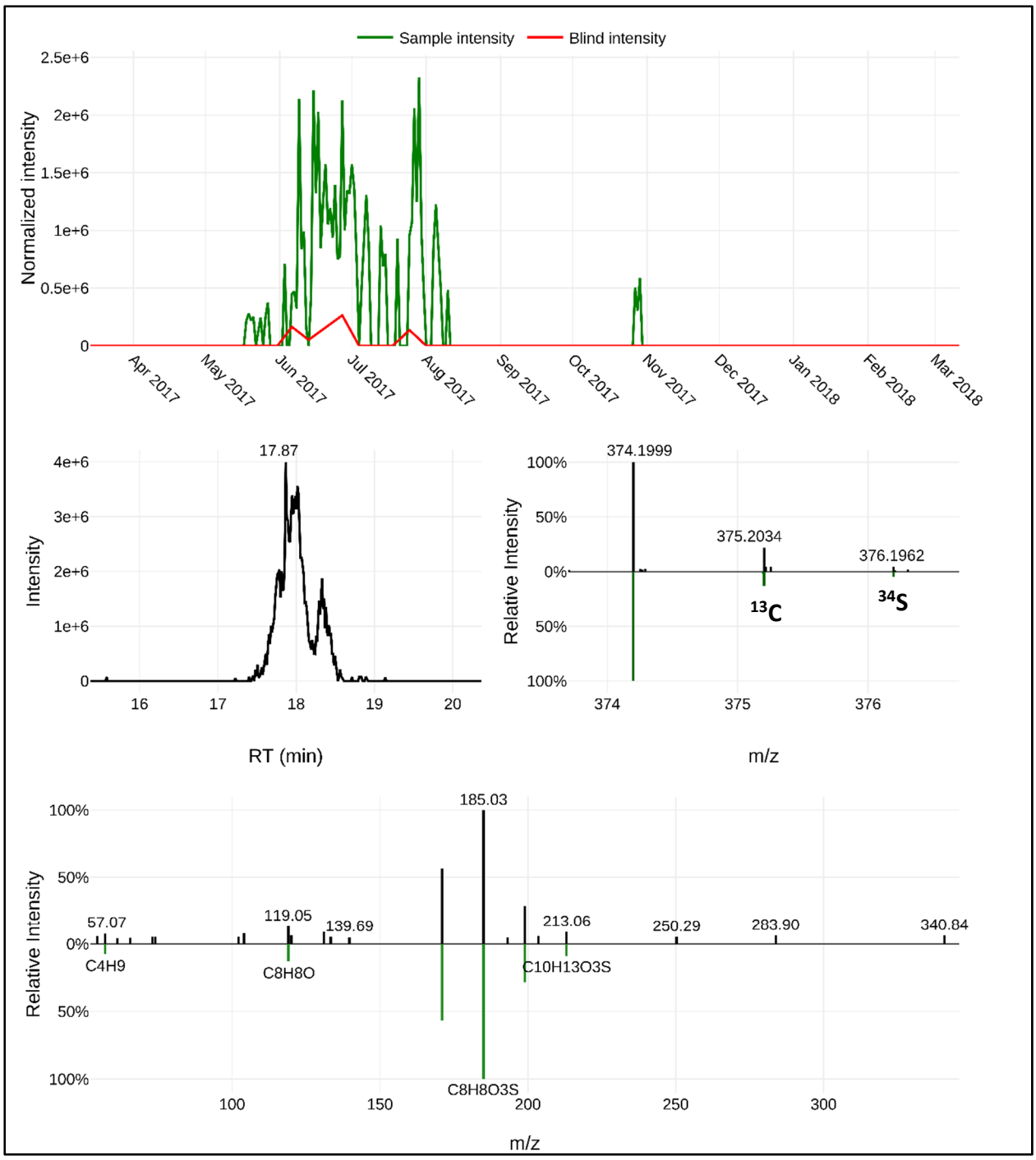



**Figure B39:** Identification of NT374 – level 2P.

NT374 likely represents 4-butoxy-N,N-bis(2-ethoxyethyl)benzenesulfonamide ($C_{18}H_{31}NO_5S$) which had a $RT_{avg}$ of 18.0 min. The proposed MS assignment matches the MS spectra ($m/z_{avg}$ = 374.1996 for $[M+H]^+$, $\Delta m$ = 0.19 ppm), the theoretical abundance (18%) of the $^{13}C$ monoisotopic mass, the theoretical abundance 4.5%) of the $^{34}S$ monoisotopic mass, and the *in silico* $MS^2$ fragments generated with Metfrag.

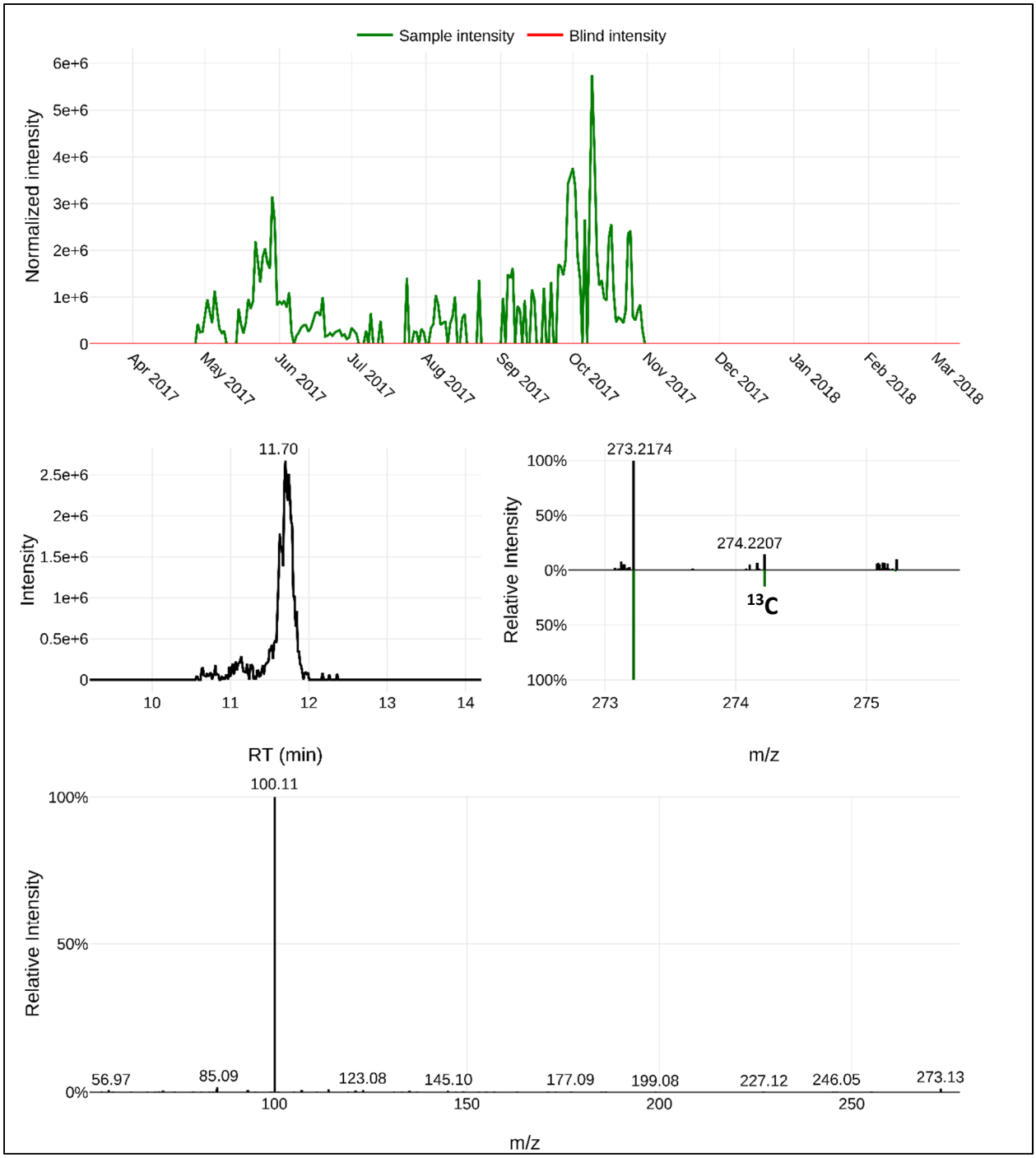



**Figure B40:** Identification of NT273 – level 4.

NT273 was identified to have an unequivocal molecular formula ($C_{14}H_{28}N_2O_3$) which had a $RT_{avg}$ of 11.6 min. The proposed molecular formula assignment matches the MS spectra ($m/z_{avg}$ = 273.2174 for $[M+H]^+$, $\Delta m$ = 0.30 ppm), the theoretical abundance (14%) of the $^{13}C$ monoisotopic mass, and the $MS^2$ fragments.

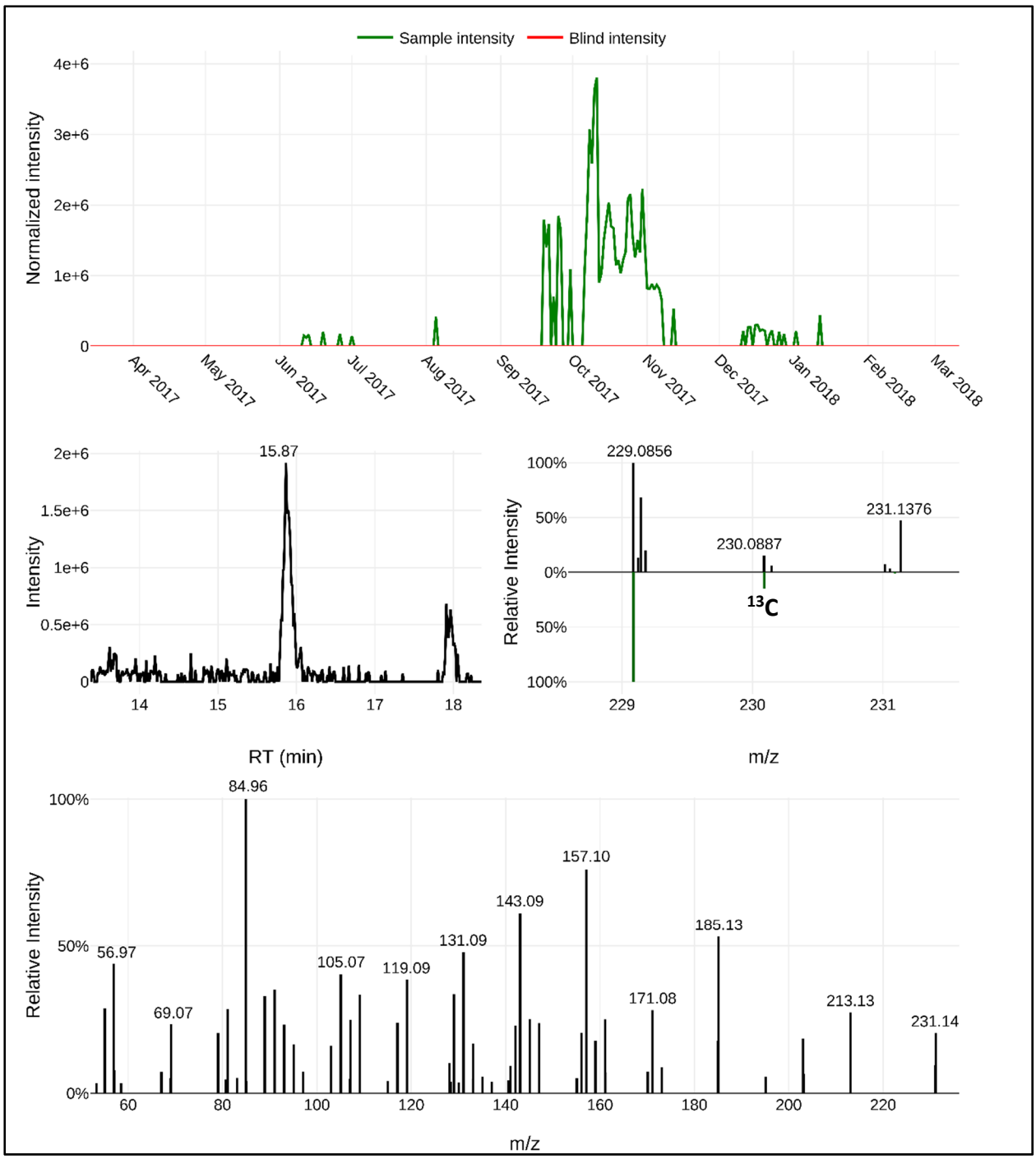



**Figure B41:** Identification of NT229 – level 4.

NT229 was identified to have an unequivocal molecular formula ($C_{14}H_{12}O_3$) which had a $RT_{avg}$ of 15.7 min. The proposed molecular formula assignment matches the MS spectra ($m/z_{avg}$ = 229.0859 for $[M+H]^+$, $\Delta m$ = -0.01 ppm), the theoretical abundance (14%) of the $^{13}C$ monoisotopic mass, and the $MS^2$ fragments.

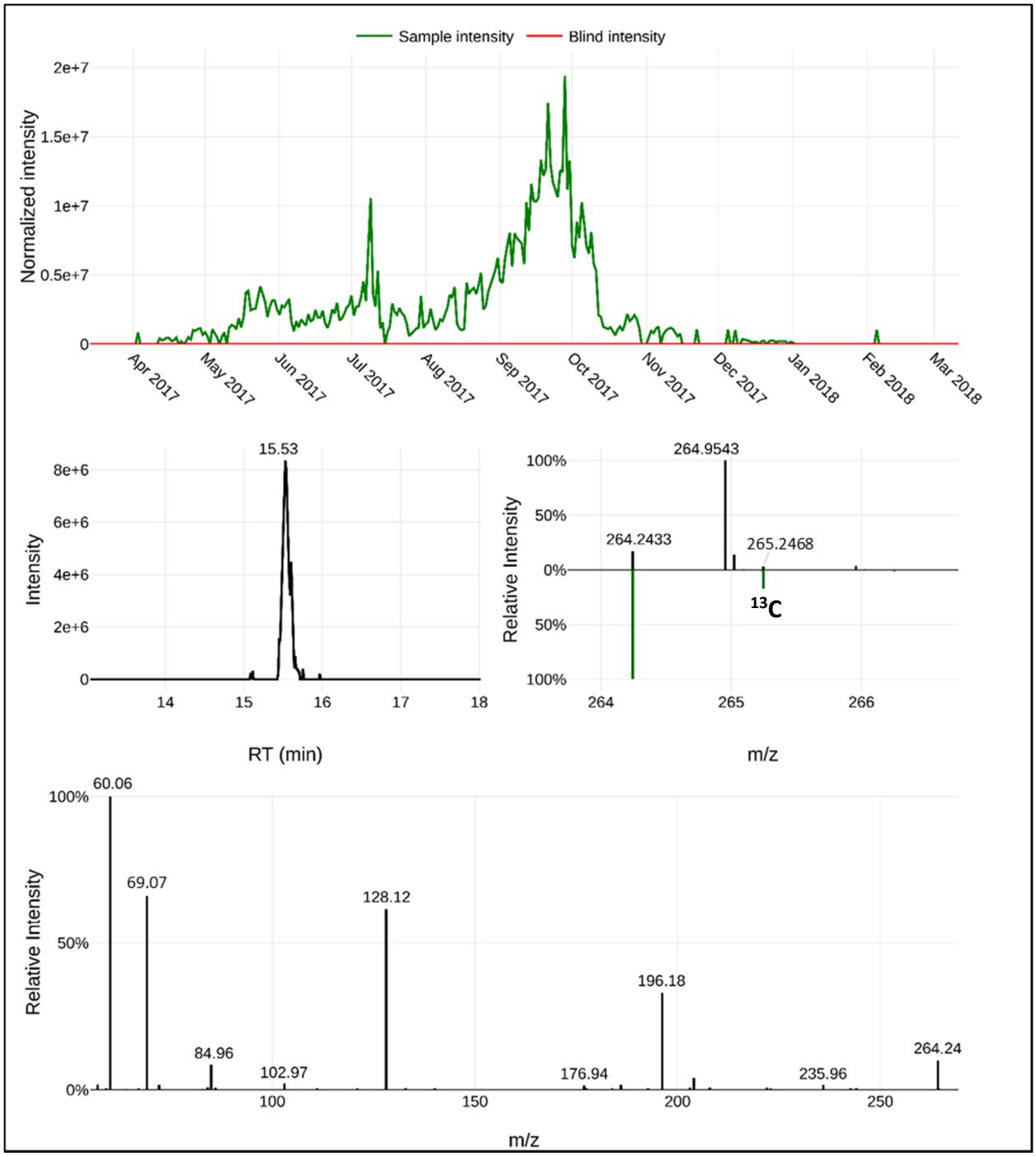



**Figure B42:** Identification of NT264 – level 3.

NT264 was identified to have an unequivocal molecular formula ($C_{16}H_{29}N_3$) which had a $RT_{avg}$ of 15.4 min. The proposed molecular formula assignment matches the MS spectra ($m/z_{avg}$ = 264.2435 for $[M+H]^+$, $\Delta m$ = 0.18 ppm), the theoretical abundance (16%) of the $^{13}C$ monoisotopic mass, and the $MS^2$ fragments. The $MS^2$ fragment ($m/z$ = 60.06) suggests a guanidine sub-structure.

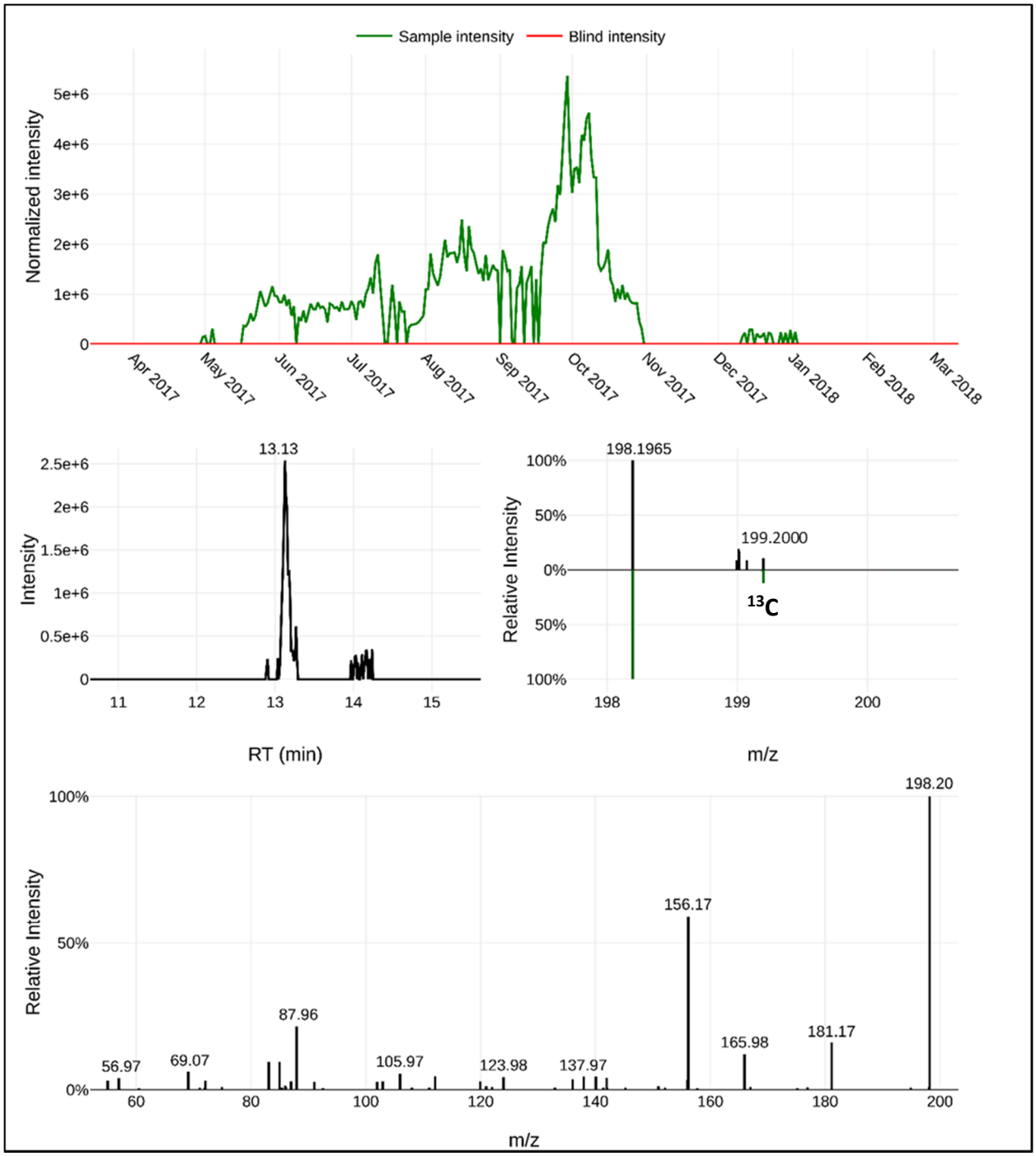



**Figure B43:** Identification of NT198 – level 4.

NT198 was identified to have an unequivocal molecular formula ($C_{11}H_{23}N_3$) which had a $RT_{avg}$ of 13.0 min. The proposed molecular formula assignment matches the MS spectra ($m/z_{avg}$ = 198.1965 for $[M+H]^+$, $\Delta m$ = 0.19 ppm), the theoretical abundance (11%) of the $^{13}C$ monoisotopic mass, and the $MS^2$ fragments.

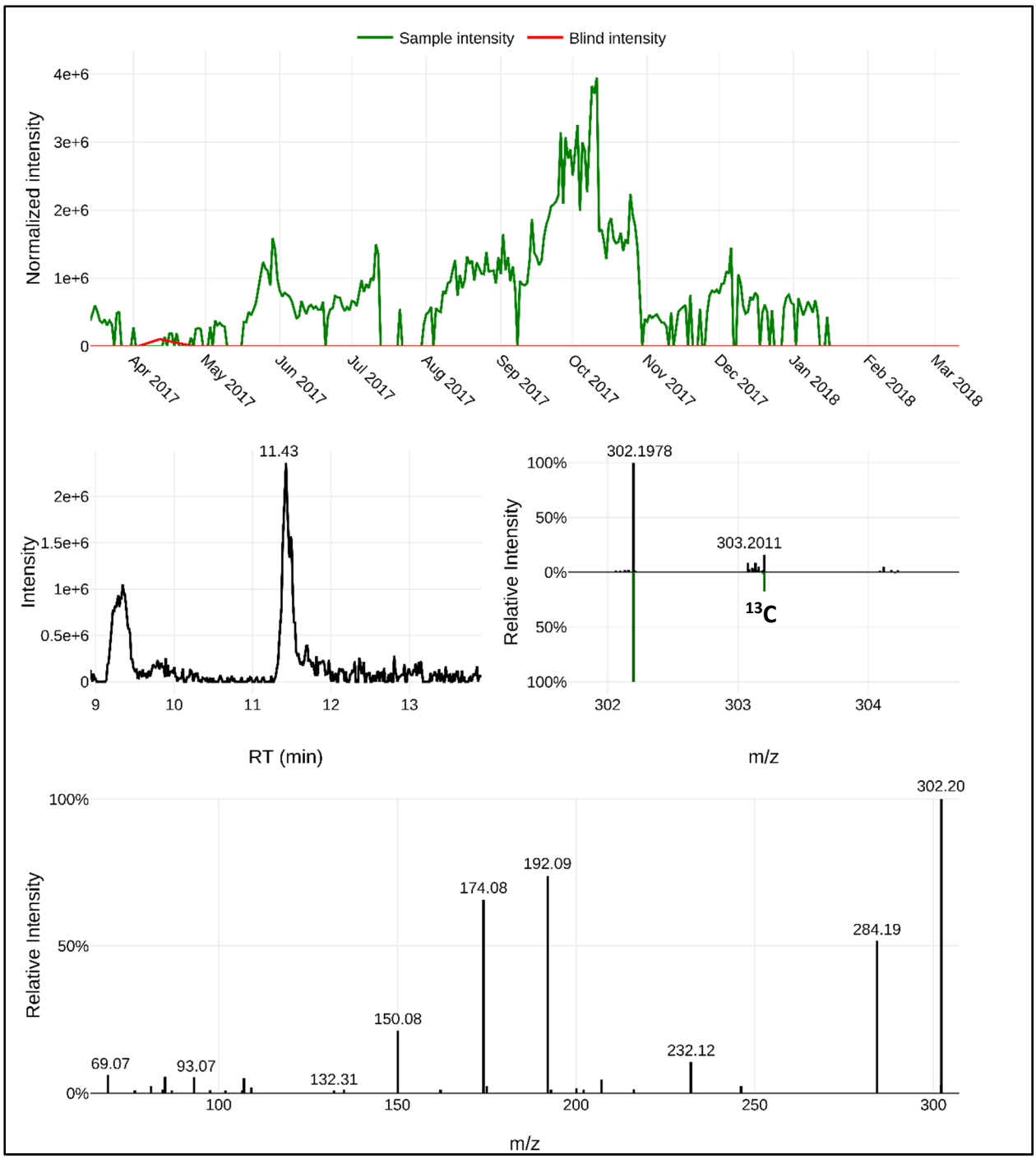



**Figure B44:** Identification of NT302 – level 4.

NT302 was identified to have an unequivocal molecular formula ($C_{16}H_{23}N_5O$) which had a $RT_{avg}$ of 11.3 min. The proposed molecular formula assignment matches the MS spectra ($m/z_{avg}$ = 302.1972 for $[M+H]^+$, $\Delta m$ = -1.07 ppm), the theoretical abundance (16%) of the $^{13}C$ monoisotopic mass, and the $MS^2$ fragments.

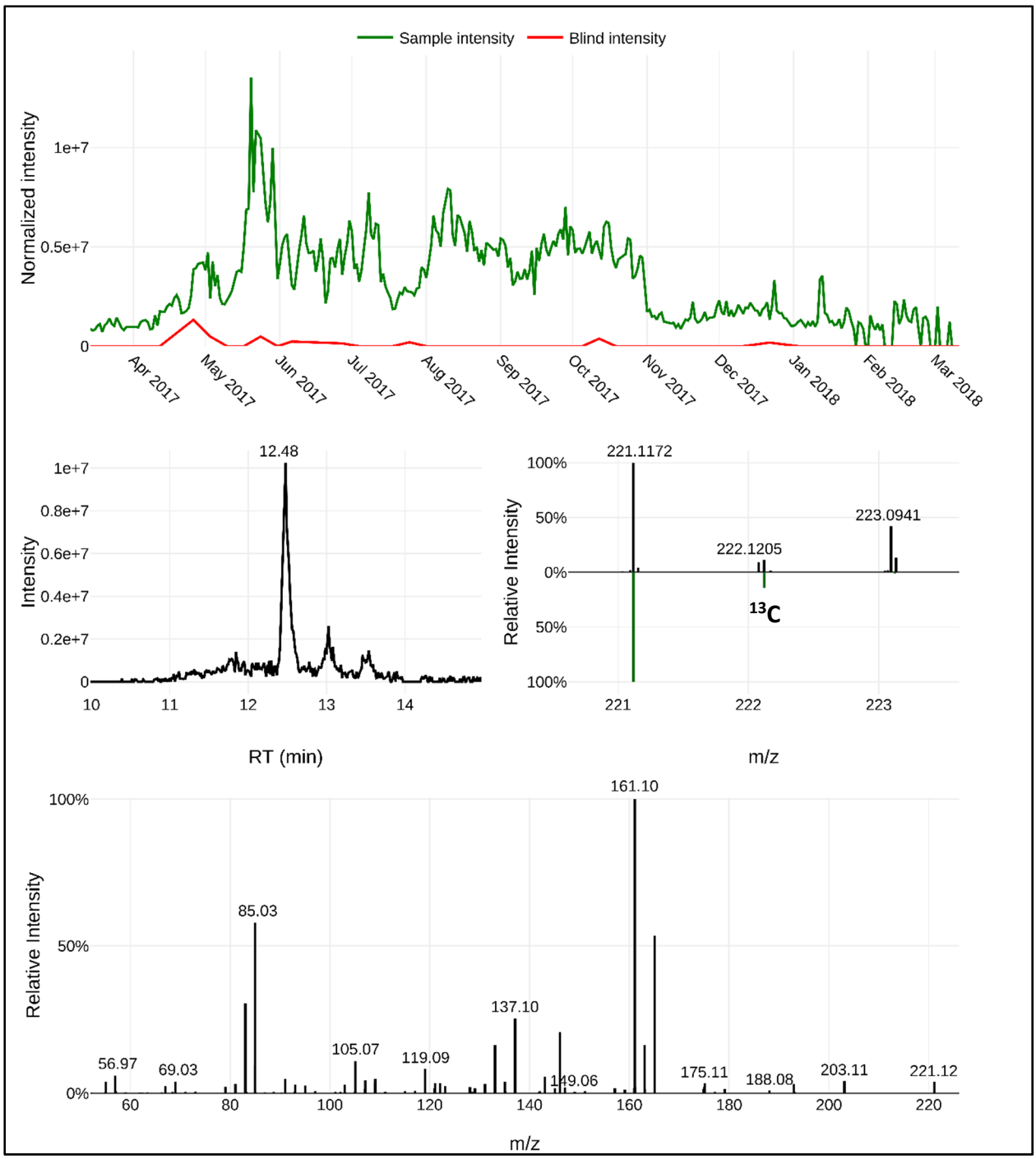



**Figure B45:** Identification of NT221 – level 4.

NT221 was identified to have an unequivocal molecular formula ($C_{13}H_{16}O_3$) which had a $RT_{avg}$ of 12.5 min. The proposed molecular formula assignment matches the MS spectra ($m/z_{avg}$ = 221.1173 for $[M+H]^+$, $\Delta m$ = 0.22 ppm), the theoretical abundance (13%) of the $^{13}C$ monoisotopic mass, and the $MS^2$ fragments.

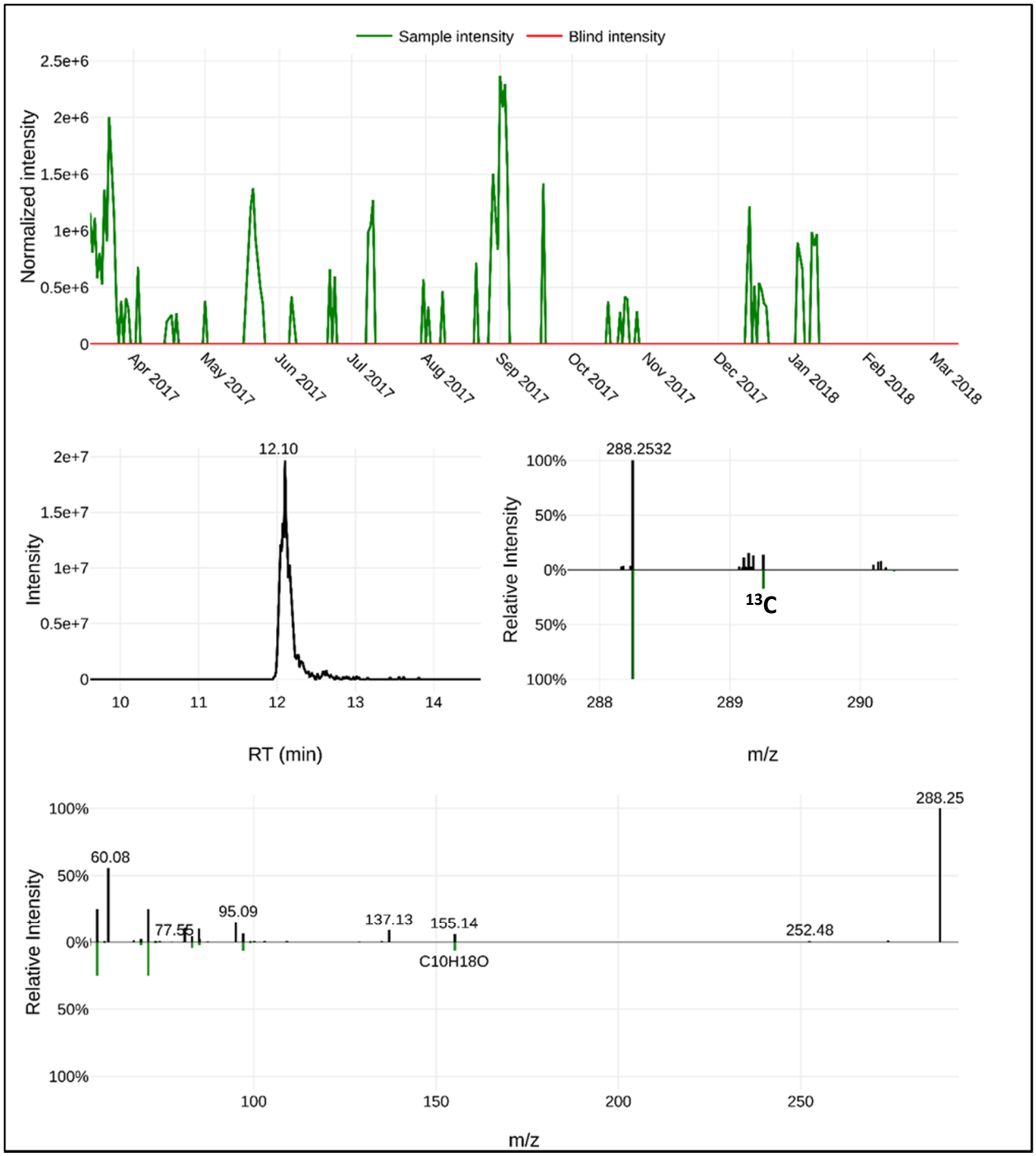



**Figure B46:** Identification of NT288 – level 2P.

NT288 likely represents lauric diethanolamide ($C_{16}H_{33}NO_3$) which had a $RT_{avg}$ of 12.1 min. The proposed MS assignment matches the MS spectra ($m/z_{avg}$ = 288.2533 for $[M+H]^+$, $\Delta m$ = 0.01 ppm), the theoretical abundance (16%) of the $^{13}C$ monoisotopic mass, and the *in silico* $MS^2$ fragments generated with Metfrag. The main $MS^2$ fragments differed from those found on MoNA; however other $MS^2$ fragments matched ($m/z$ = 55.05, 55.07, 67.05, 69.07, 81.07, 83.08, 95.09, and 97.10).

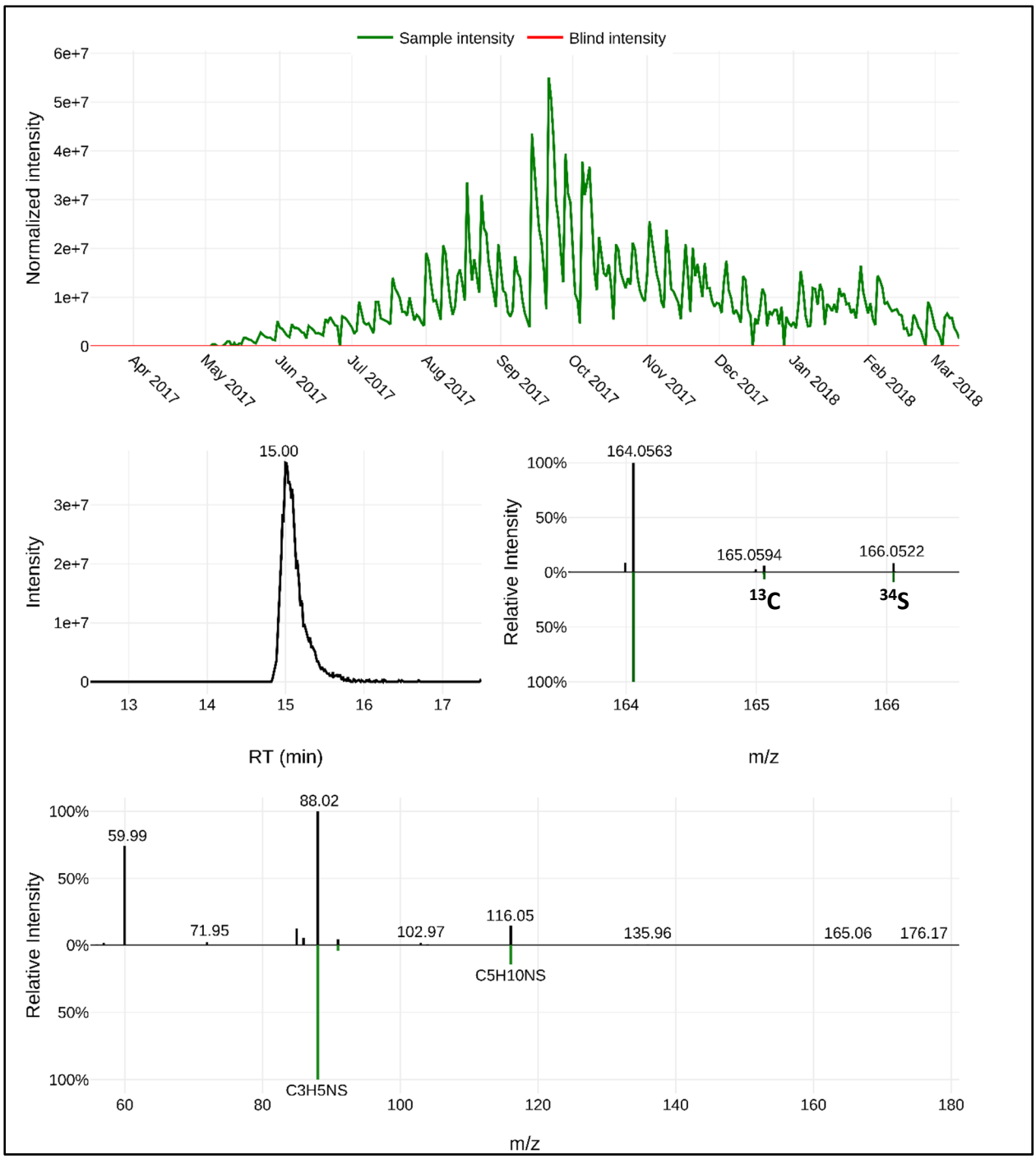



**Figure B47:** Identification of methyl diethyldithiocarbamate – level 1.

Methyl diethyldithiocarbamate ($C_6H_{13}NS_2$) was confirmed using an authentic reference standard which matched the $RT_{avg}$ of 14.9 min, the MS spectra ($m/z_{avg}$ = 164.0561 for $[M+H]^+$, $\Delta m$ = -0.69 ppm), the theoretical abundance (6%) of the $^{13}C$ monoisotopic mass, the theoretical abundance (9%) of the $^{34}S$ monoisotopic mass, and the $MS^2$ fragments ($m/z$ = 59.99, 88.02, and 116.05).

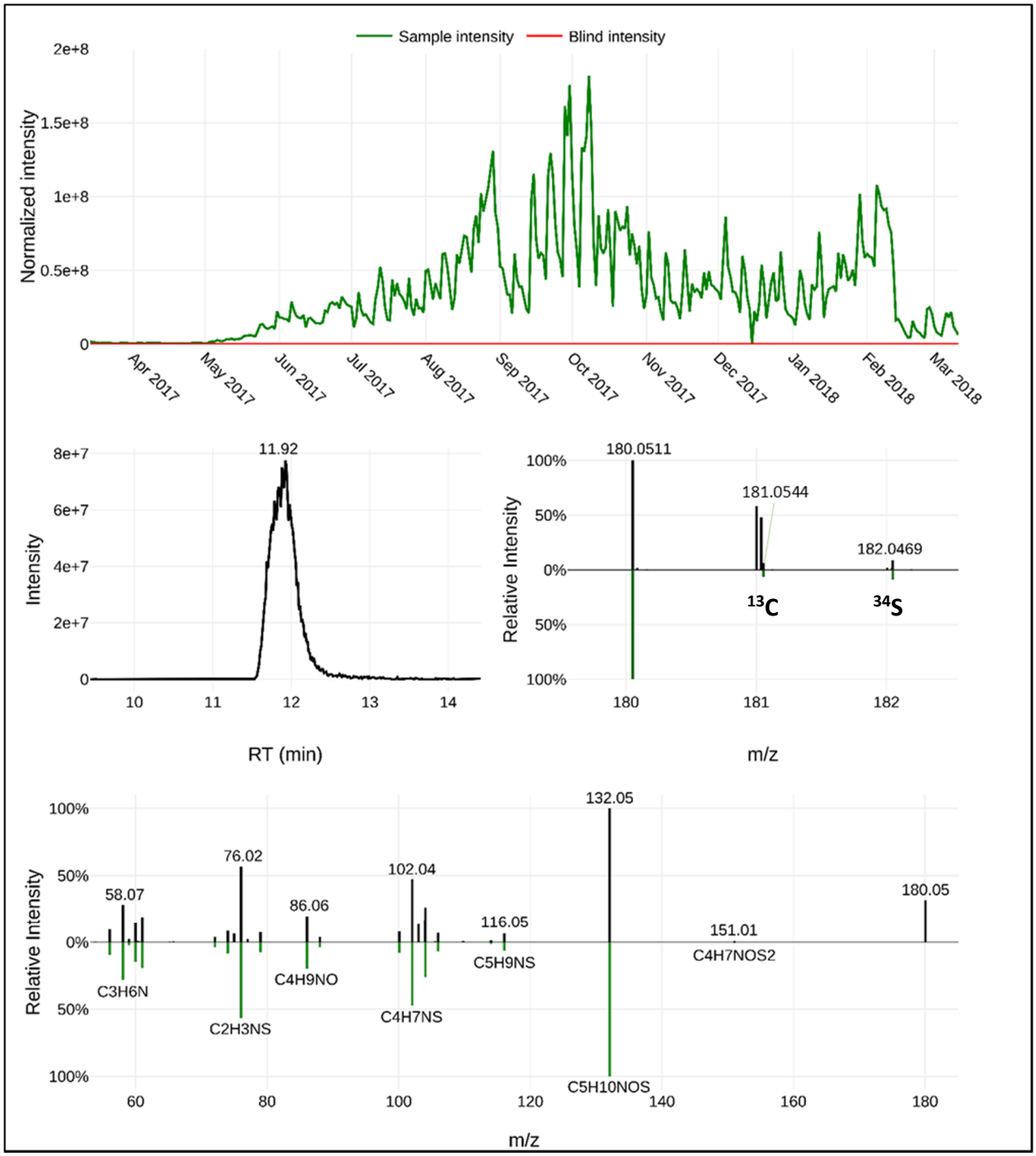



**Figure B48:** Identification of NT180 – level 3.

NT180 was identified to have an unequivocal molecular formula ($C_6H_{13}NOS_2$) which had a $RT_{avg}$ of 11.7 min. The proposed molecular formula assignment matches the MS spectra ($m/z_{avg}$ = 180.0511 for $[M+H]^+$, $\Delta m$ = -0.01 ppm), the theoretical abundance (6%) of the $^{13}C$ monoisotopic mass, the theoretical abundance (9%) of the $^{34}S$ monoisotopic mass, and the $MS^2$ fragments. The $MS^2$ fragments ($m/z$ = 88.02) suggest a thiocarbamate sub-structure.

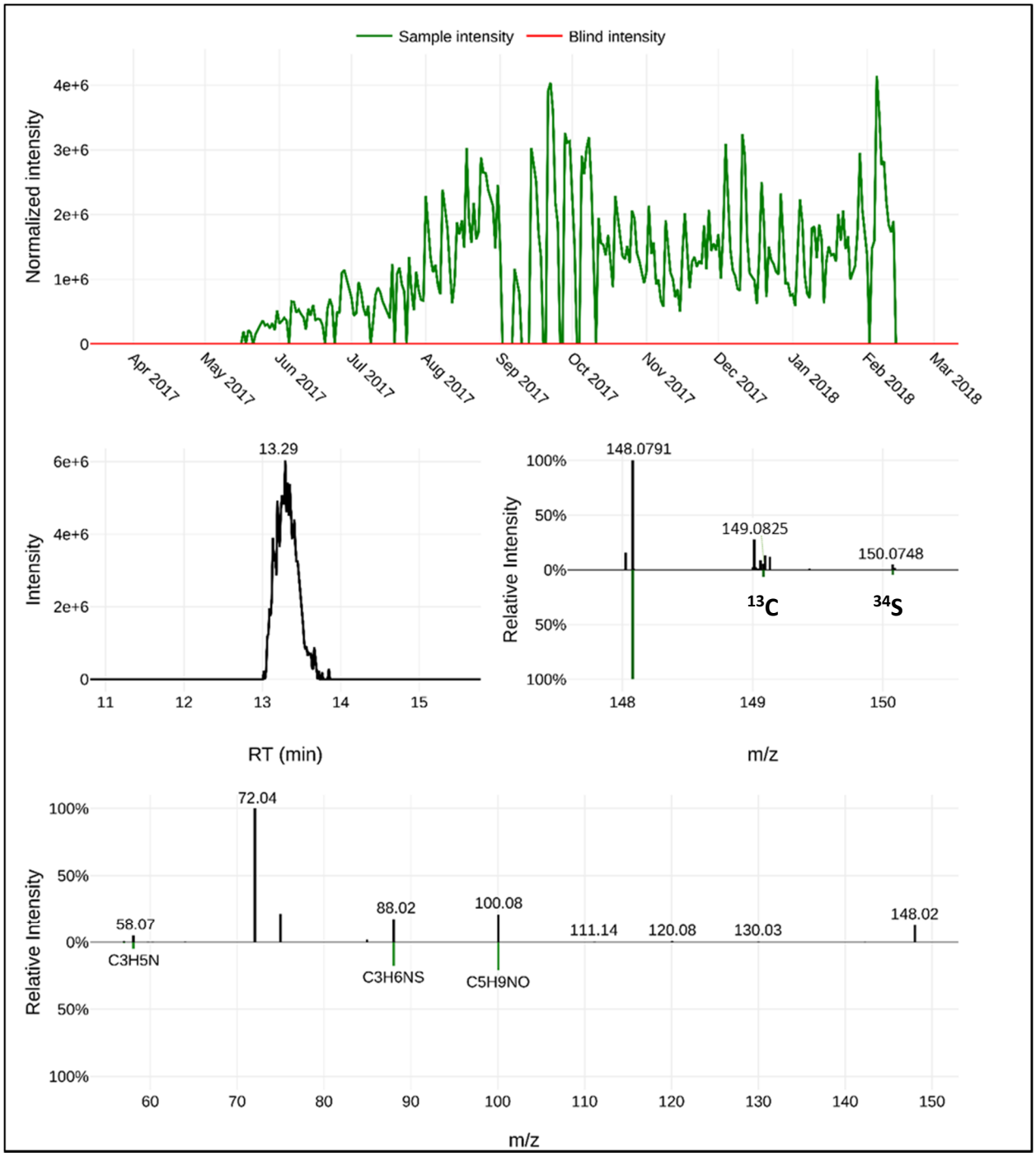



**Figure B49:** Identification of NT148 – level 3.

NT148 was identified to have an unequivocal molecular formula ($C_6H_{13}NOS$) which had a $RT_{avg}$ of 13.3 min. The proposed molecular formula assignment matches the MS spectra ($m/z_{avg}$ = 148.0790 for $[M+H]^+$, $\Delta m$ = -0.51 ppm), the theoretical abundance (6%) of the $^{13}C$ monoisotopic mass, the theoretical abundance (4.5%) of the $^{34}S$ monoisotopic mass, and the $MS^2$ fragments. The $MS^2$ fragments ($m/z$ = 88.02) suggest a thiocarbamate sub-structure.

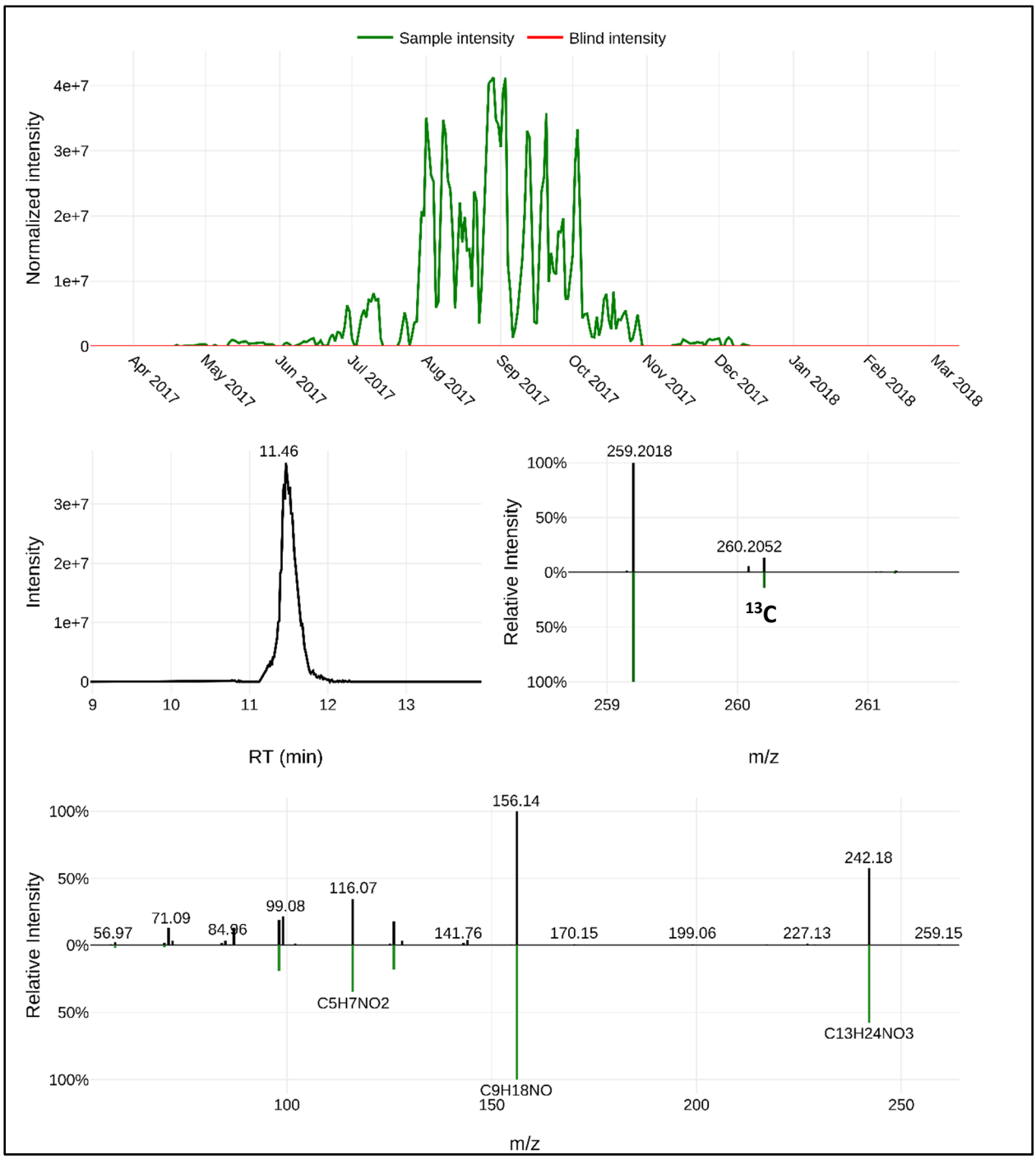



**Figure B50:** Identification of NT259 – level 2P.

NT259 likely represents (4S)-4-Amino-5-(dibutylamino)-5-oxopentanoic acid ($C_{13}H_{26}N_2O_3$) which had a $RT_{avg}$ of 11.4 min. The proposed MS assignment matches the MS spectra ($m/z_{avg}$ = 259.2017 for $[M+H]^+$, $\Delta m$ = -0.40 ppm), the theoretical abundance (13%) of the $^{13}C$ monoisotopic mass, and the *in silico* $MS^2$ fragments generated with Metfrag.

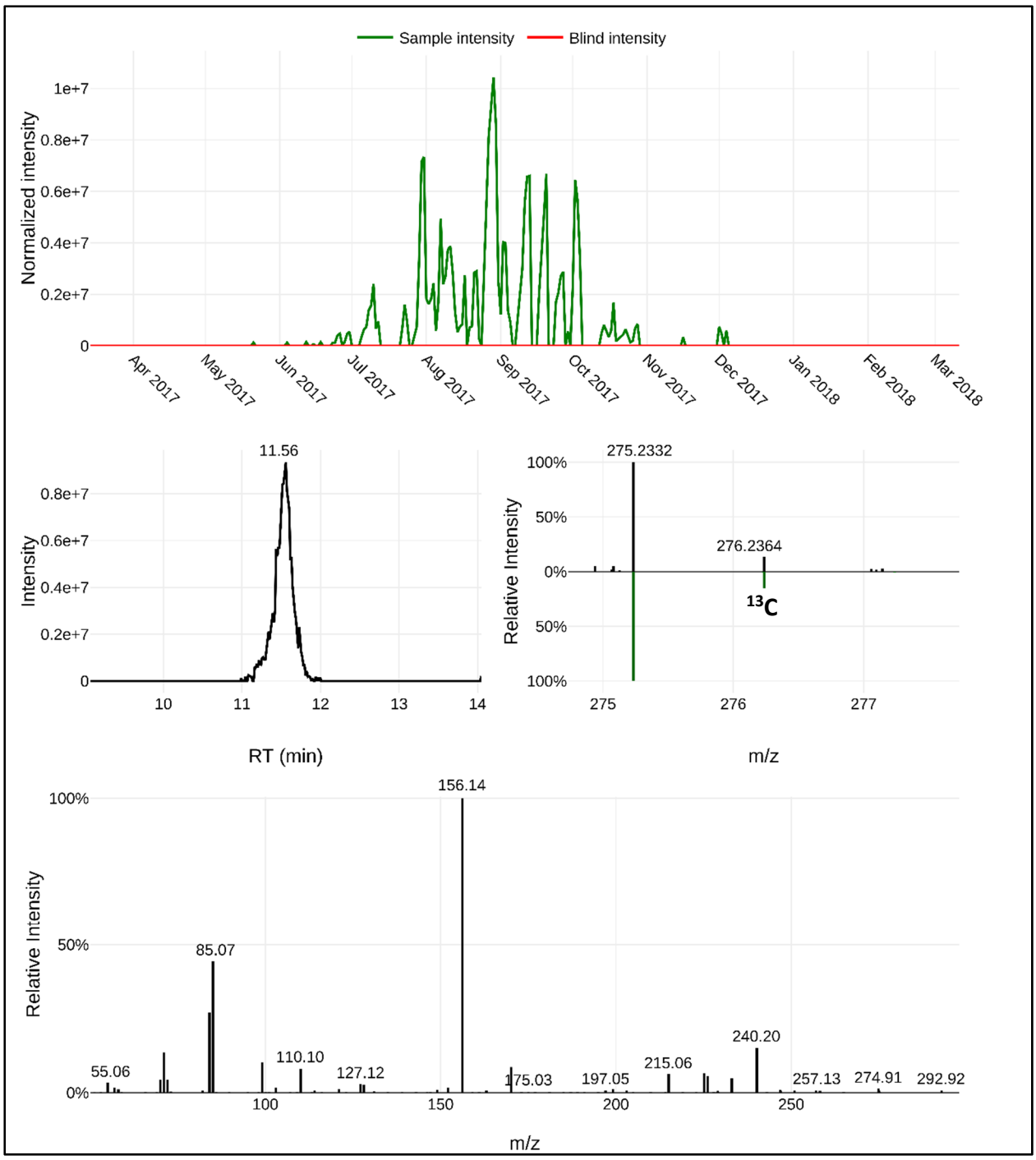



**Figure B51:** Identification of NT275 – level 4.

NT275 was identified to have an unequivocal molecular formula ($C_{14}H_{30}N_2O_3$) which had a $RT_{avg}$ of 11.5 min. The proposed molecular formula assignment matches the MS spectra ($m/z_{avg}$ = 275.2330 for $[M+H]^+$, $\Delta m$ = 0.45 ppm), the theoretical abundance (14%) of the $^{13}C$ monoisotopic mass, and the $MS^2$ fragments.

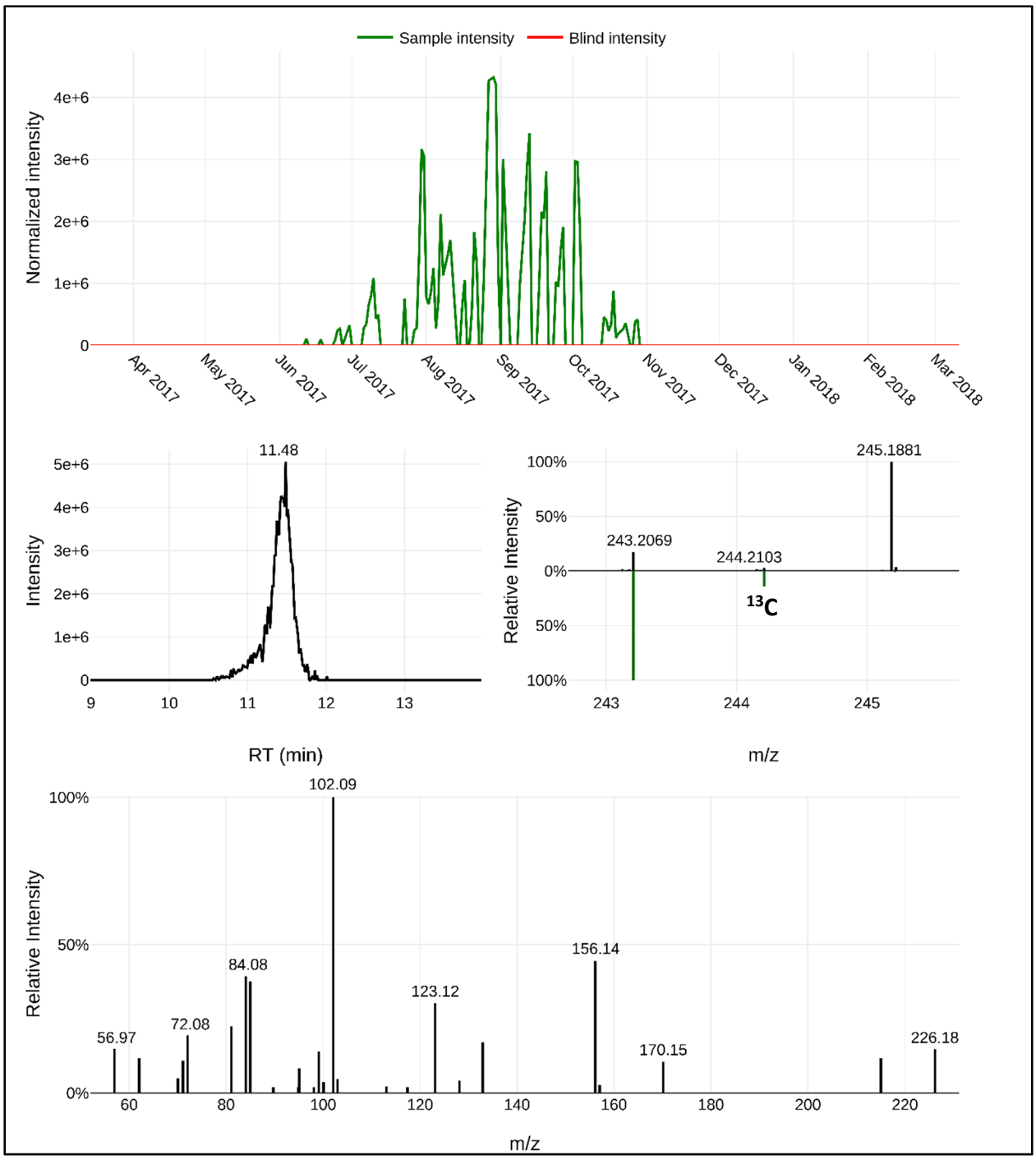



**Figure B52:** Identification of NT243 – level 4.

NT243 was identified to have an unequivocal molecular formula ($C_{13}H_{26}N_2O_2$) which had a $RT_{avg}$ of 11.6 min. The proposed molecular formula assignment matches the MS spectra ($m/z_{avg}$ = 243.2068 for $[M+H]^+$, $\Delta m$ = 0.36 ppm), the theoretical abundance (13%) of the $^{13}C$ monoisotopic mass, and the $MS^2$ fragments.

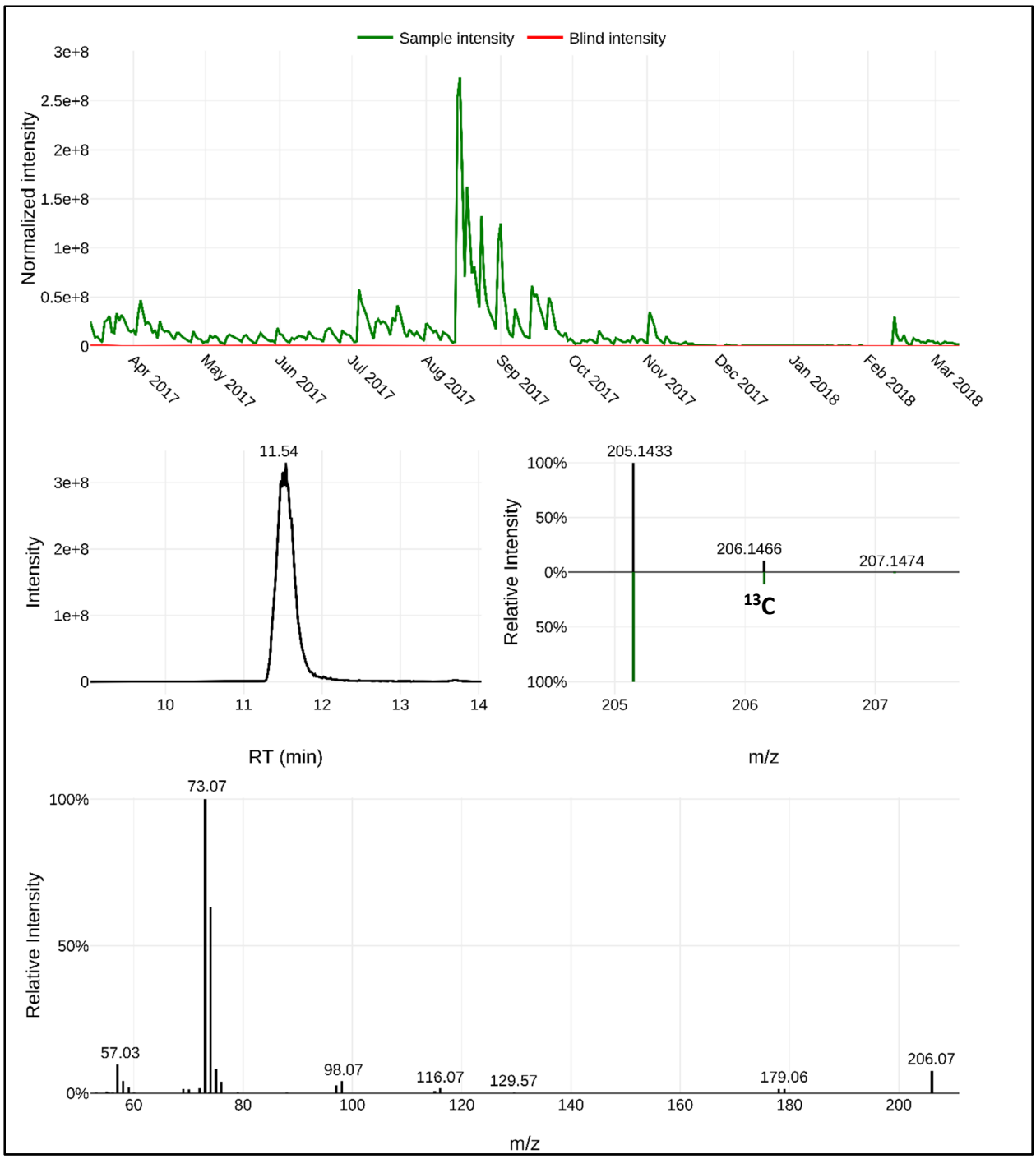



**Figure B53:** Identification of NT205a – level 4.

NT205a was identified to have an unequivocal molecular formula ($C_{10}H_{20}O_4$) which had a $RT_{avg}$ of 11.6 min. The proposed molecular formula assignment matches the MS spectra ($m/z_{avg}$ = 205.1435 for $[M+H]^+$, $\Delta m$ = 0.43 ppm), the theoretical abundance (10%) of the $^{13}C$ monoisotopic mass, and the $MS^2$ fragments.

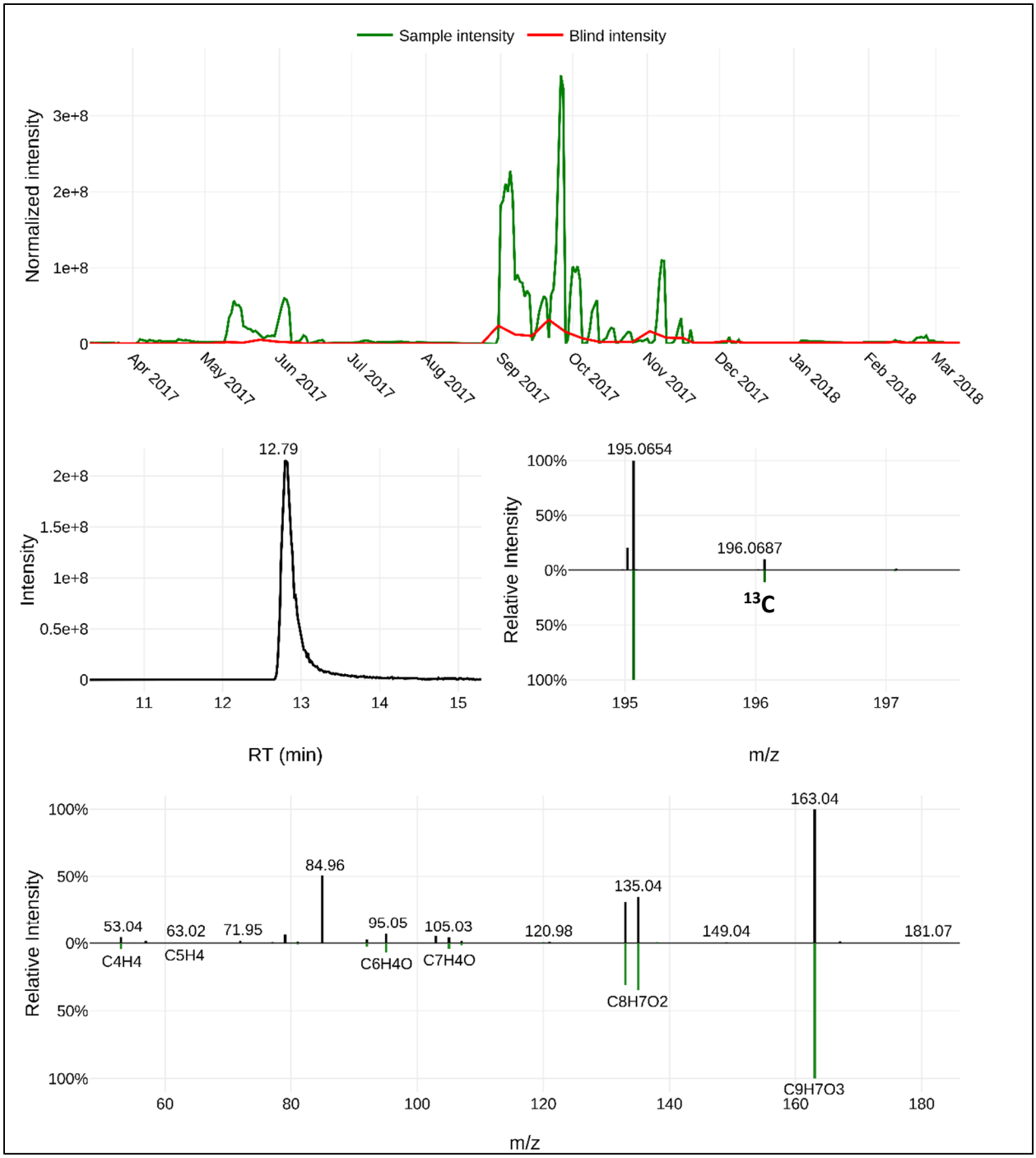



**Figure B54:** Identification of dimethyl phthalate – level 1.

Dimethyl phthalate ($C_{10}H_{10}O_4$) was confirmed using an authentic reference standard which matched the $RT_{avg}$ of 12.7 min, the MS spectra ($m/z_{avg}$ = 195.0652 for $[M+H]^+$, $\Delta m$ = 0.31 ppm), the theoretical abundance (10%) of the $^{13}C$ monoisotopic mass, and the $MS^2$ fragments ($m/z$ = 133.03, 135.04, and 163.04).

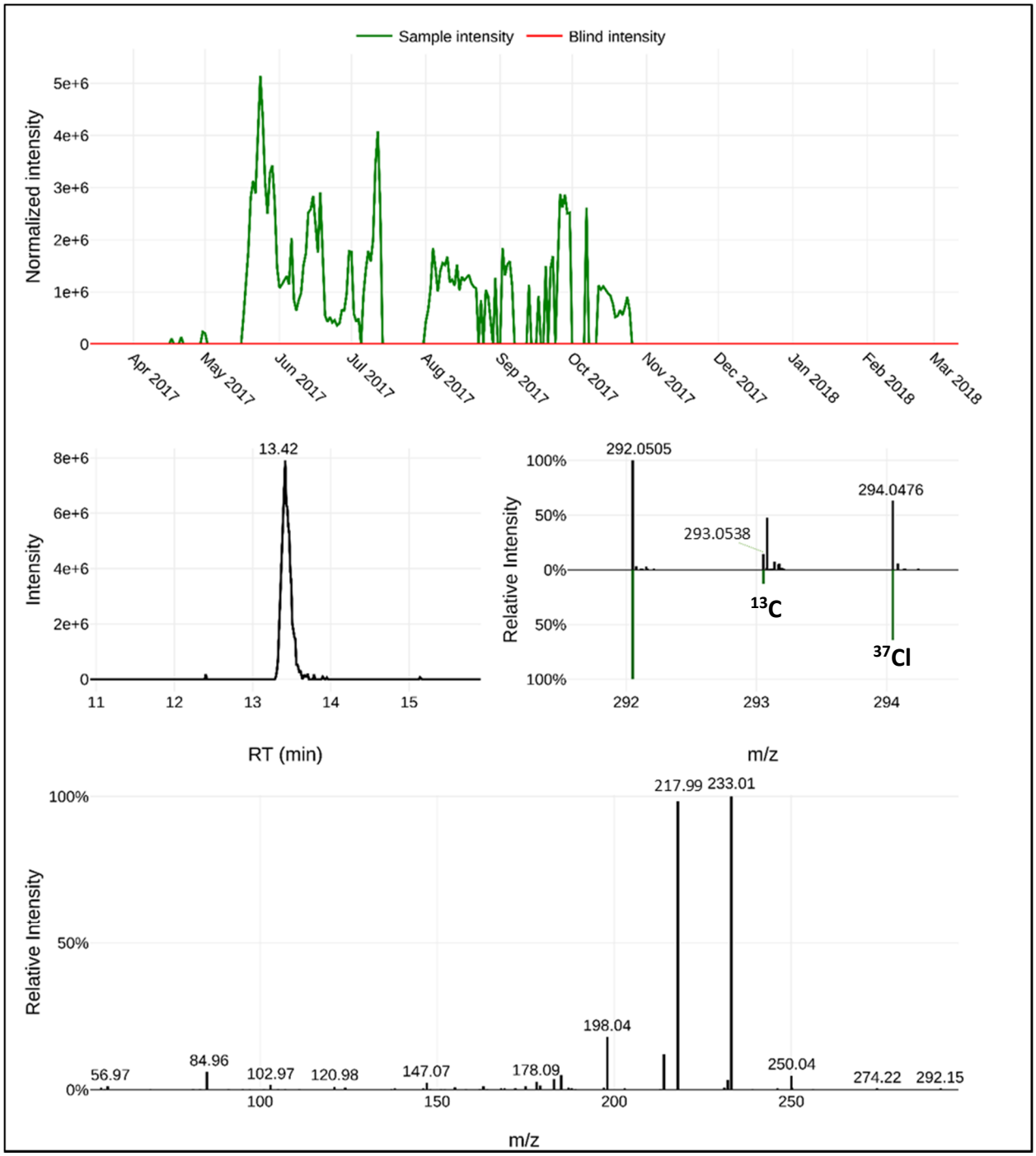



**Figure B55:** Identification of NT292 – level 4.

NT292 was identified to have an unequivocal molecular formula ($C_{12}H_{15}NO_3Cl_2$) which had a $RT_{avg}$ of 13.5 min. The proposed molecular formula assignment matches the MS spectra ($m/z_{avg}$ = 292.0502 for $[M+H]^+$, $\Delta m$ = 0.14 ppm), the theoretical abundance (12%) of the $^{13}C$ monoisotopic mass, the theoretical abundance (64%) of the $^{37}Cl$ monoisotopic mass, and the $MS^2$ fragments.

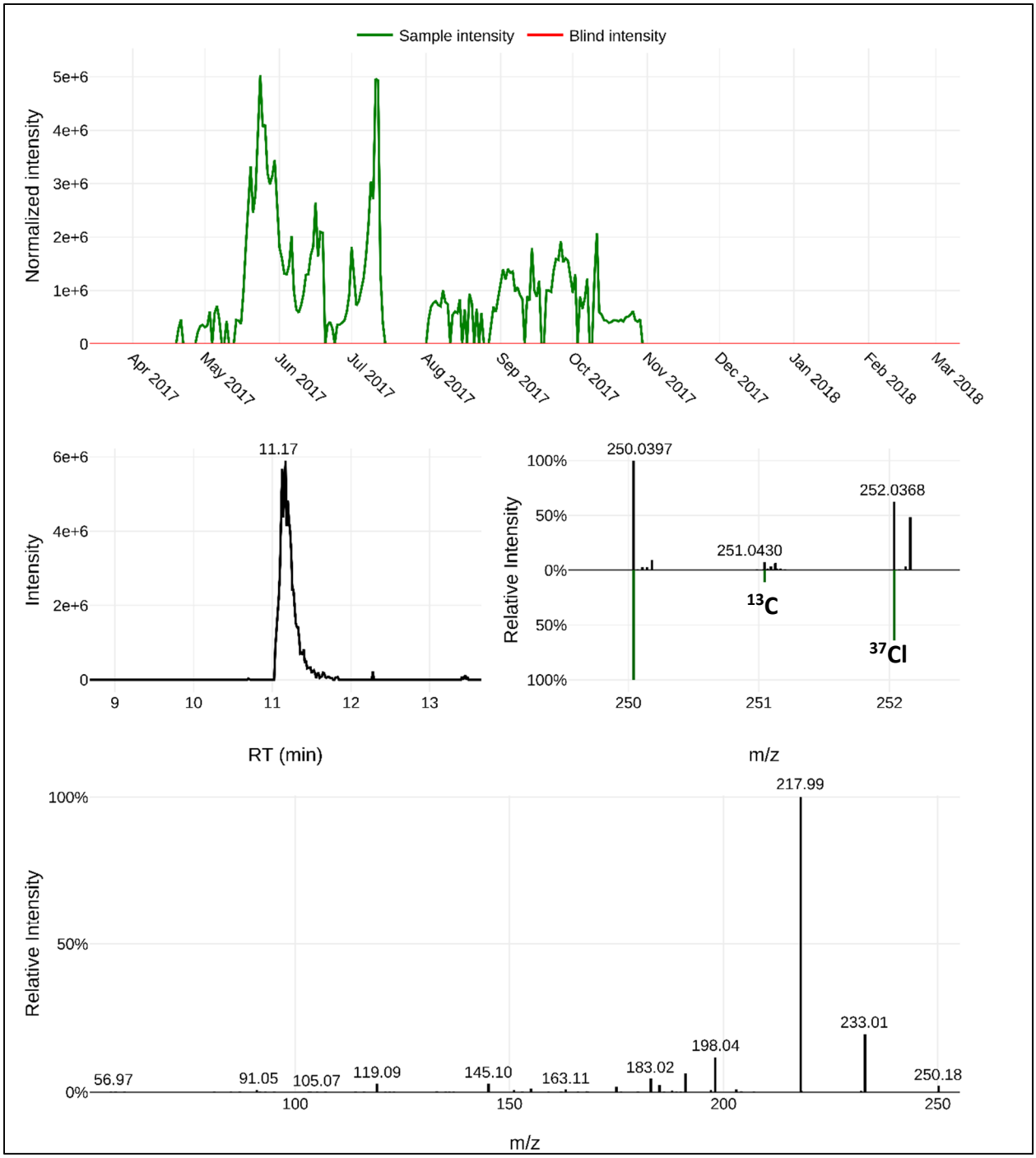



**Figure B56:** Identification of NT250 – level 4.

NT250 was identified to have an unequivocal molecular formula ($C_{10}H_{13}NO_2Cl_2$) which had a $RT_{avg}$ of 11.2 min. The proposed molecular formula assignment matches the MS spectra ($m/z_{avg}$ = 250.0396 for $[M+H]^+$, $\Delta m$ = 0.04 ppm), the theoretical abundance (10%) of the $^{13}C$ monoisotopic mass, the theoretical abundance (64%) of the $^{37}Cl$ monoisotopic mass, and the $MS^2$ fragments.

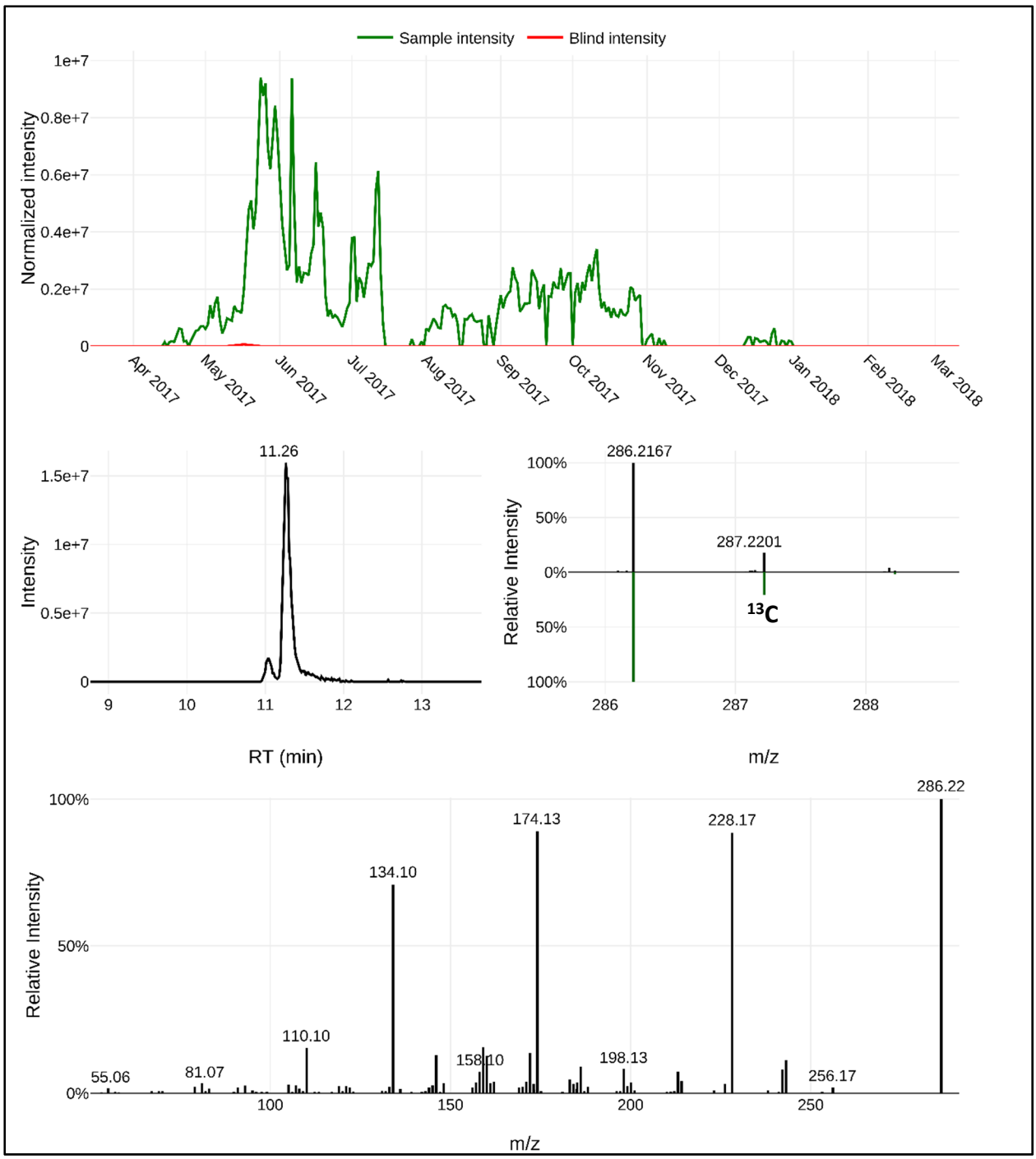



**Figure B57:** Identification of NT286 – level 4.

NT286 was identified to have an unequivocal molecular formula ($C_{19}H_{27}NO$) which had a $RT_{avg}$ of 11.3 min. The proposed molecular formula assignment matches the MS spectra ($m/z_{avg}$ = 286.2166 for $[M+H]^+$, $\Delta m$ = 0.24 ppm), the theoretical abundance (19%) of the $^{13}C$ monoisotopic mass, and the $MS^2$ fragments.

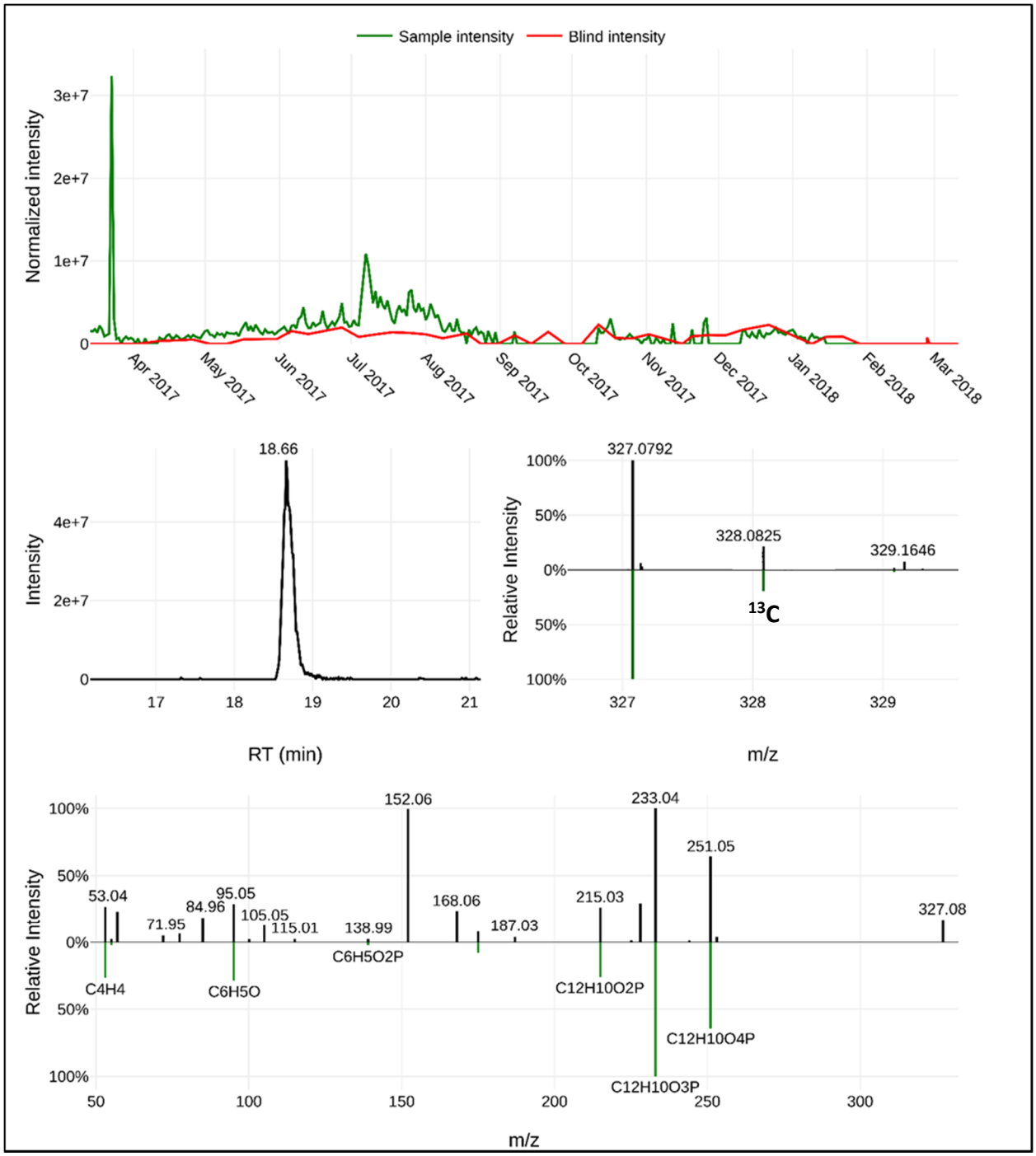



**Figure B58:** Identification of triphenyl phosphate – level 1.

Triphenyl phosphate ($C_{18}H_{15}O_4P$) was confirmed using an authentic reference standard which matched the $RT_{avg}$ of 18.7 min, the MS spectra ($m/z_{avg}$ = 327.0781 for $[M+H]^+$, $\Delta m$ = 0.00 ppm), the theoretical abundance (18%) of the $^{13}C$ monoisotopic mass, and the $MS^2$ fragments ($m/z$ = 152.06, 233.04, and 251.05).

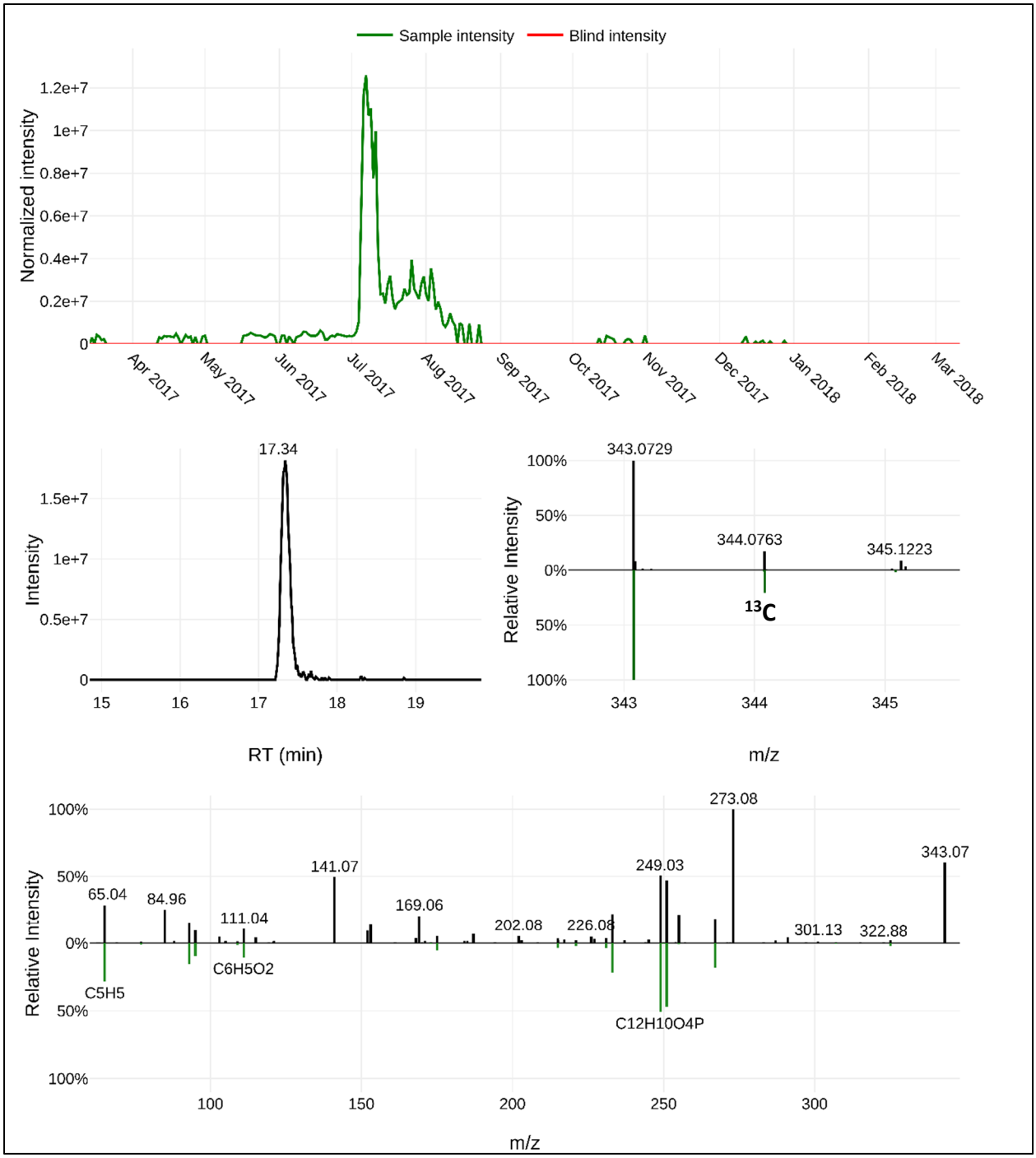



**Figure B59:** Identification of NT343 – level 3.

NT343 was identified to have an unequivocal molecular formula ($C_{18}H_{15}O_5P$) which had a $RT_{avg}$ of 17.4 min. The proposed molecular formula assignment matches the MS spectra ($m/z_{avg}$ = 343.0731 for $[M+H]^+$, $\Delta m$ = 0.26 ppm), the theoretical abundance (18%) of the $^{13}C$ monoisotopic mass, and the $MS^2$ fragments. The $MS^2$ fragments ($m/z$ = 233.04 and 251.05) suggest a triphenyl phosphate sub-structure. NT343 likely represents hydroxyphenyl diphenyl phosphate.

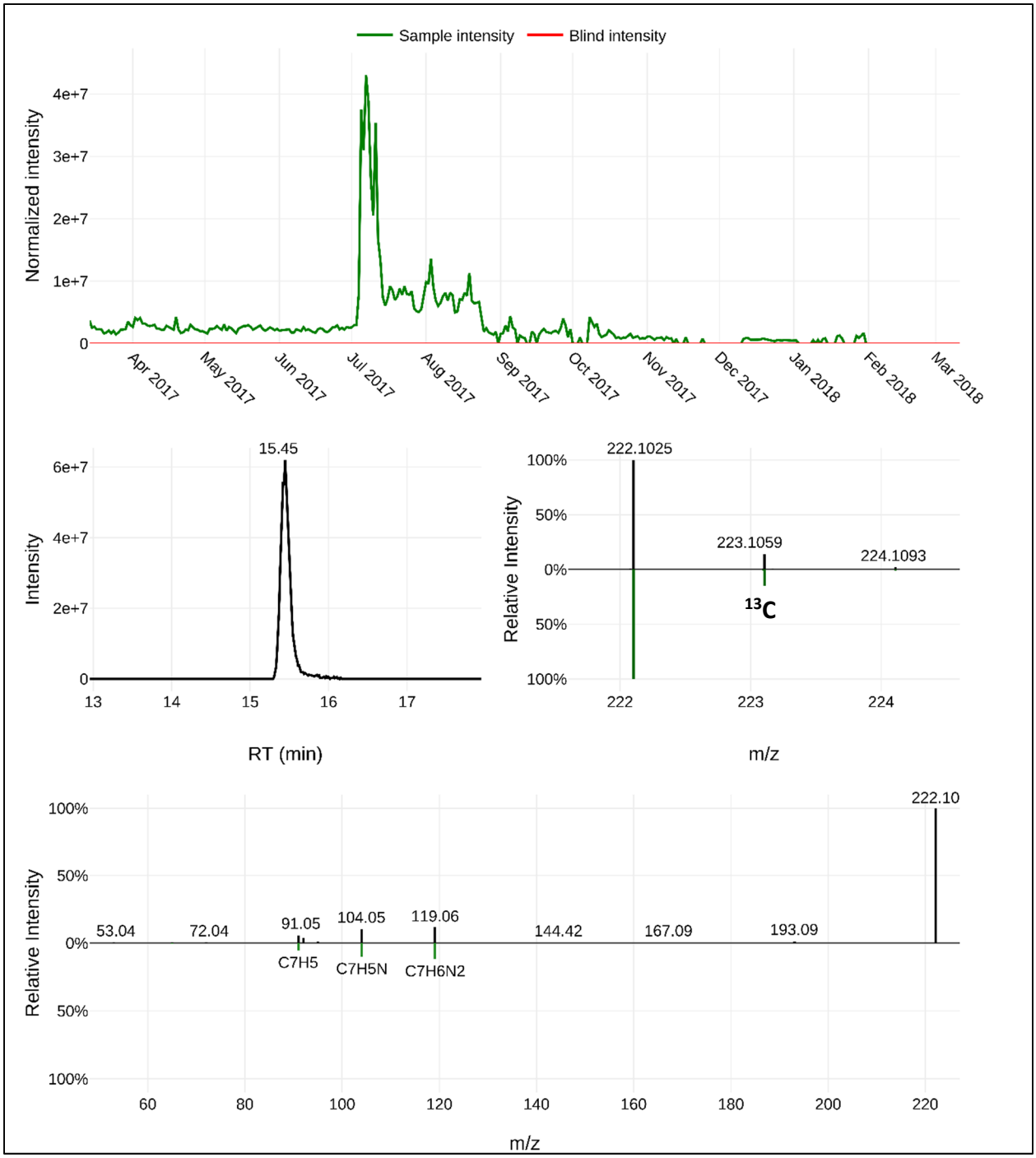



**Figure B60:** Identification of NT222 – level 3.

NT222 was identified to have an unequivocal molecular formula ($C_{14}H_{11}N_3$) which had a $RT_{avg}$ of 15.5 min. The proposed molecular formula assignment matches the MS spectra ($m/z_{avg}$ = 222.1026 for $[M+H]^+$, $\Delta m$ = 0.03 ppm), the theoretical abundance (14%) of the $^{13}C$ monoisotopic mass, and the $MS^2$ fragments. The $MS^2$ fragments ($m/z$ = 92.05 and 119.06) suggest a quinazoline sub-structure. NT222 likely represents anilinoquinazoline.

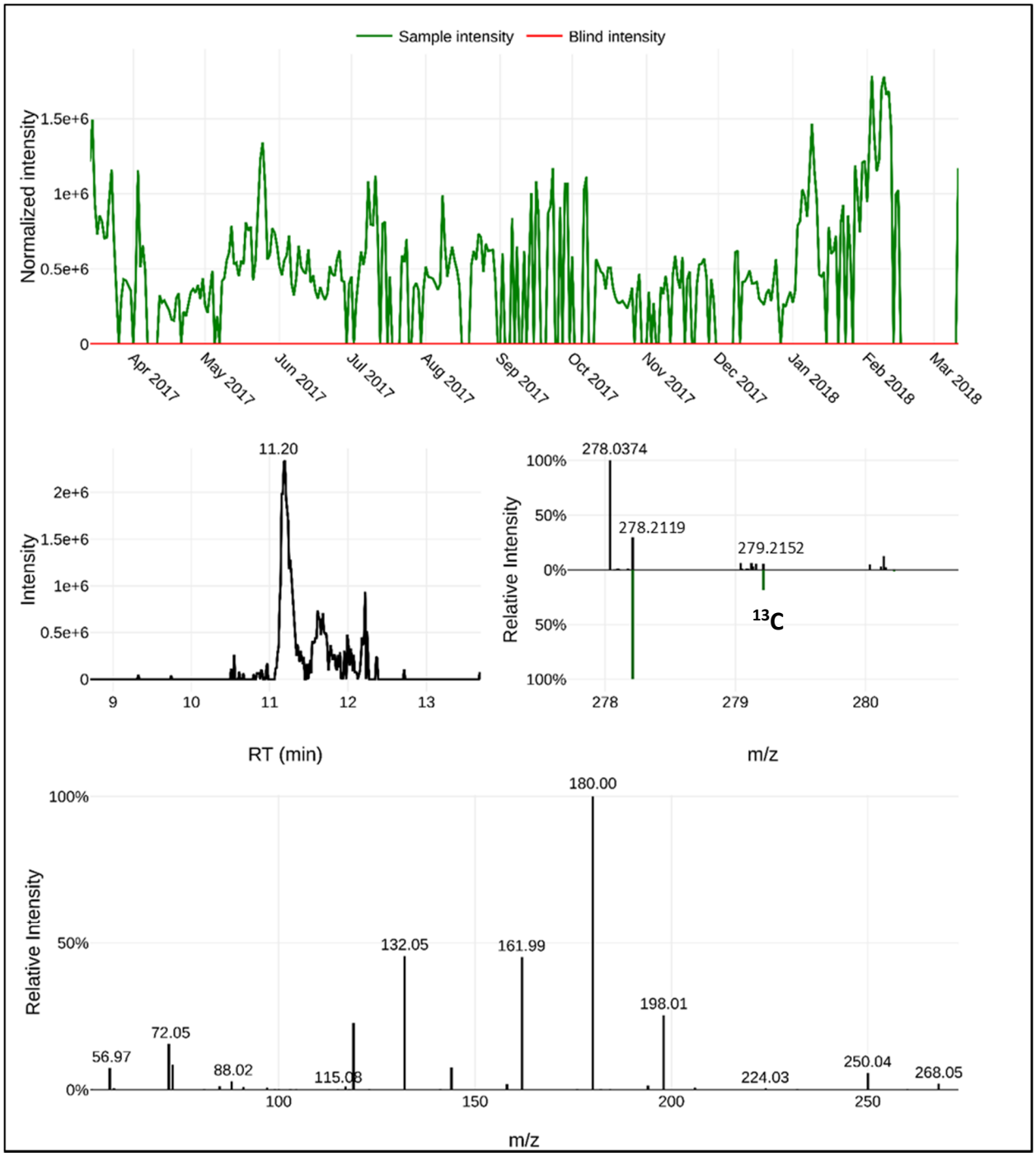



**Figure B61:** Identification of NT278a – level 4.

NT278a was identified to have an unequivocal molecular formula ($C_{17}H_{27}NO_2$) which had a $RT_{avg}$ of 11.4 min. The proposed molecular formula assignment matches the MS spectra ($m/z_{avg}$ = 278.2116 for $[M+H]^+$, $\Delta m$ = 0.36 ppm), the theoretical abundance (17%) of the $^{13}C$ monoisotopic mass, and the $MS^2$ fragments.

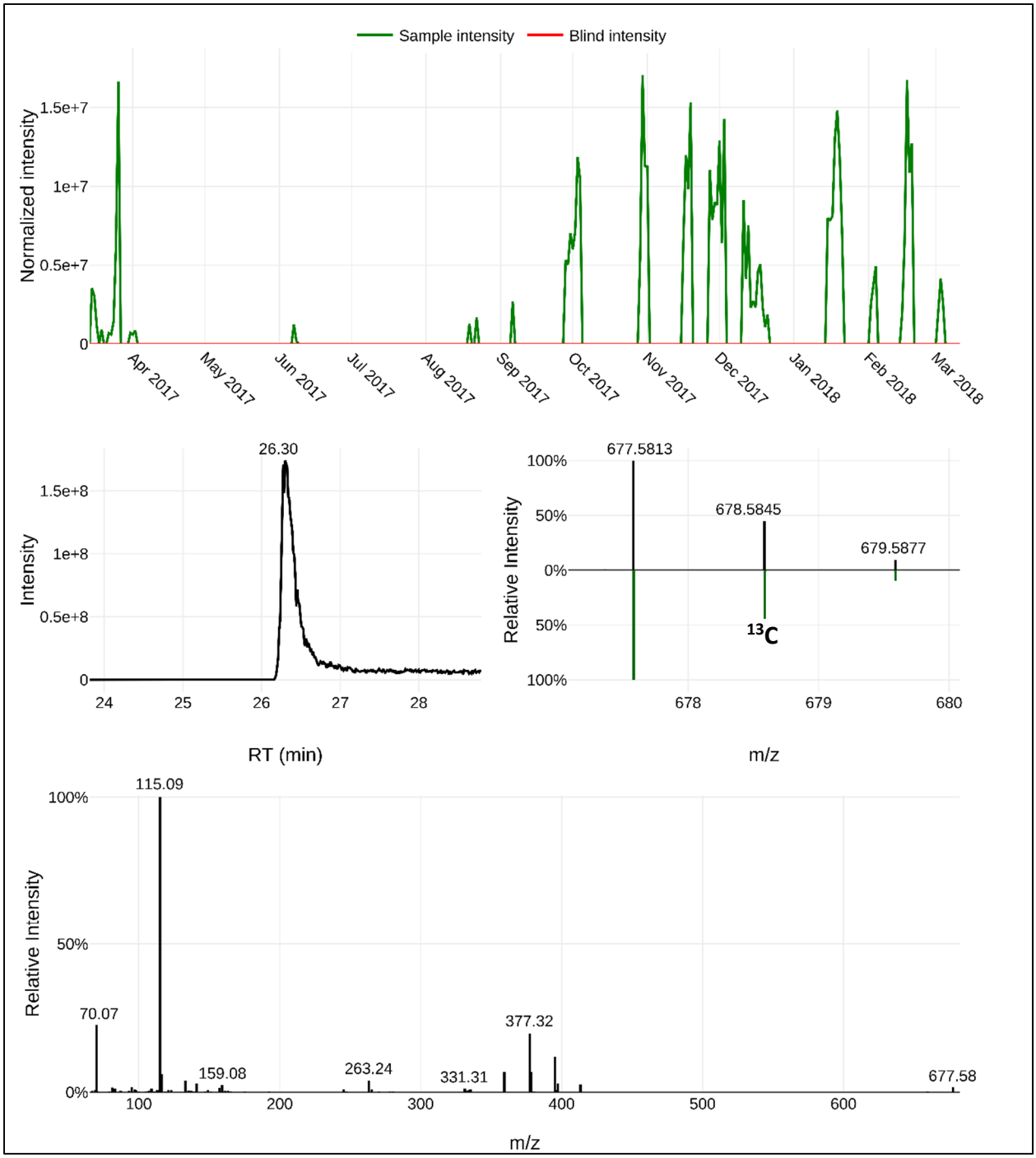



**Figure B62:** Identification of NT677 – level 4.

NT677 was identified to have an unequivocal molecular formula ($C_{41}H_{76}N_2O_5$) which had a $RT_{avg}$ of 26.9 min. The proposed molecular formula assignment matches the MS spectra ($m/z_{avg}$ = 677.5821 for $[M+H]^+$, $\Delta m$ = -0.96 ppm), the theoretical abundance (41%) of the $^{13}C$ monoisotopic mass, and the $MS^2$ fragments.

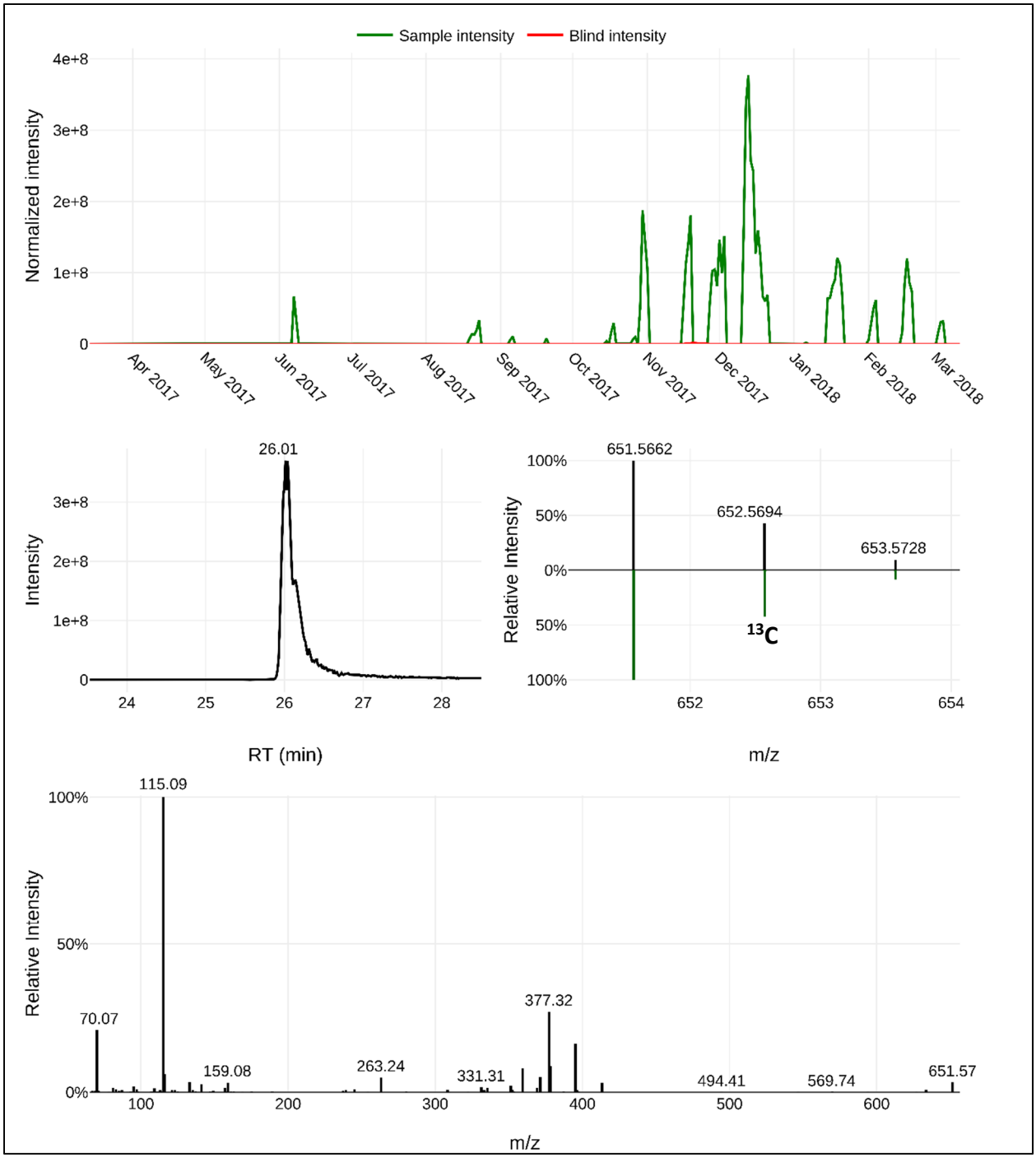



**Figure B63:** Identification of NT651 – level 4.

NT651 was identified to have an unequivocal molecular formula ($C_{39}H_{74}N_2O_5$) which had a $RT_{avg}$ of 26.2 min. The proposed molecular formula assignment matches the MS spectra ($m/z_{avg}$ = 651.5665 for $[M+H]^+$, $\Delta m$ = -0.87 ppm), the theoretical abundance (39%) of the $^{13}C$ monoisotopic mass, and the $MS^2$ fragments.

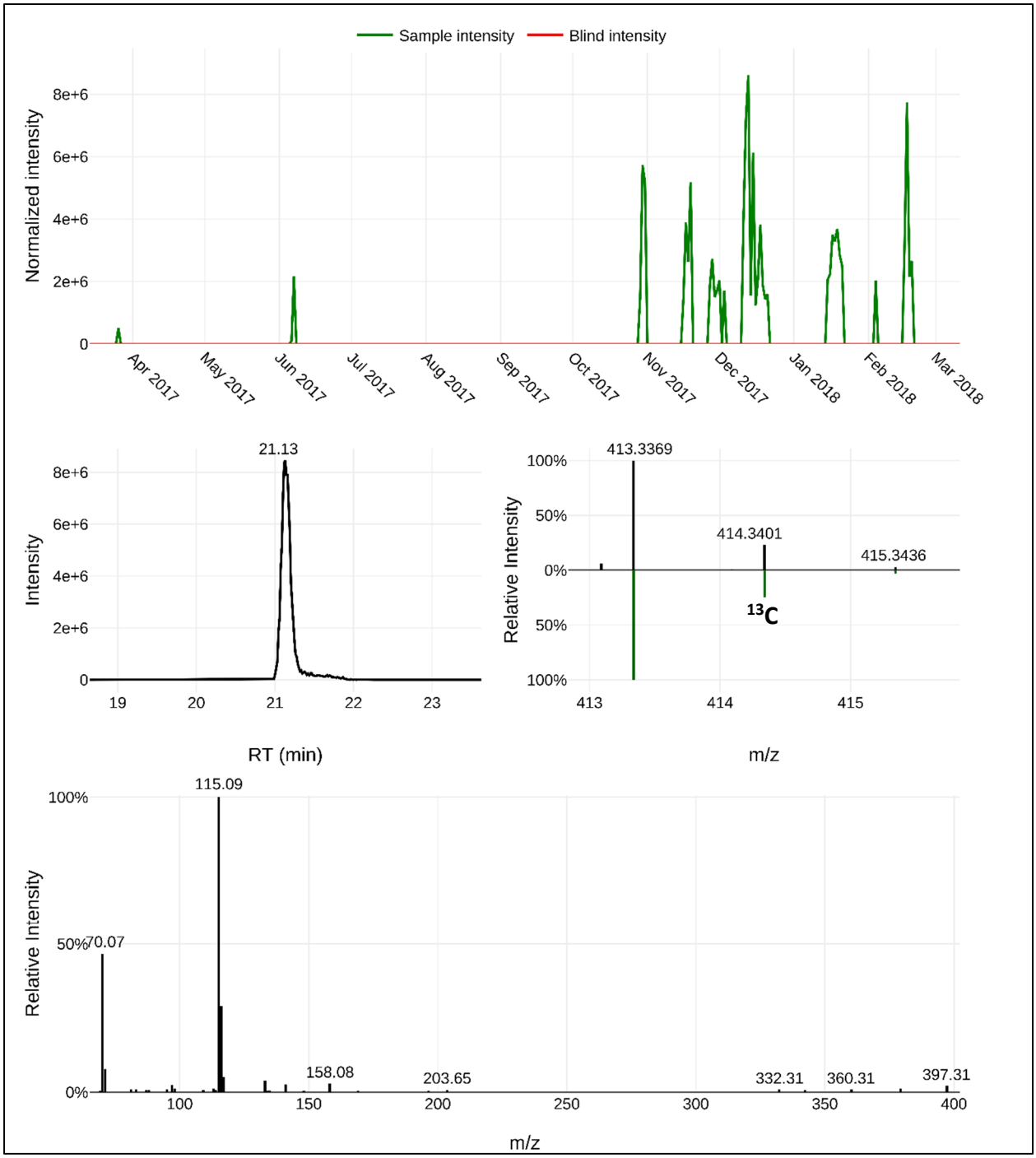



**Figure B64:** Identification of NT413 – level 4.

NT413 was identified to have an unequivocal molecular formula ($C_{23}H_{44}N_2O_4$) which had a $RT_{avg}$ of 21.2 min. The proposed molecular formula assignment matches the MS spectra ($m/z_{avg}$ = 413.3372 for $[M+H]^+$, $\Delta m$ = -0.33 ppm), the theoretical abundance (23%) of the $^{13}C$ monoisotopic mass, and the $MS^2$ fragments.

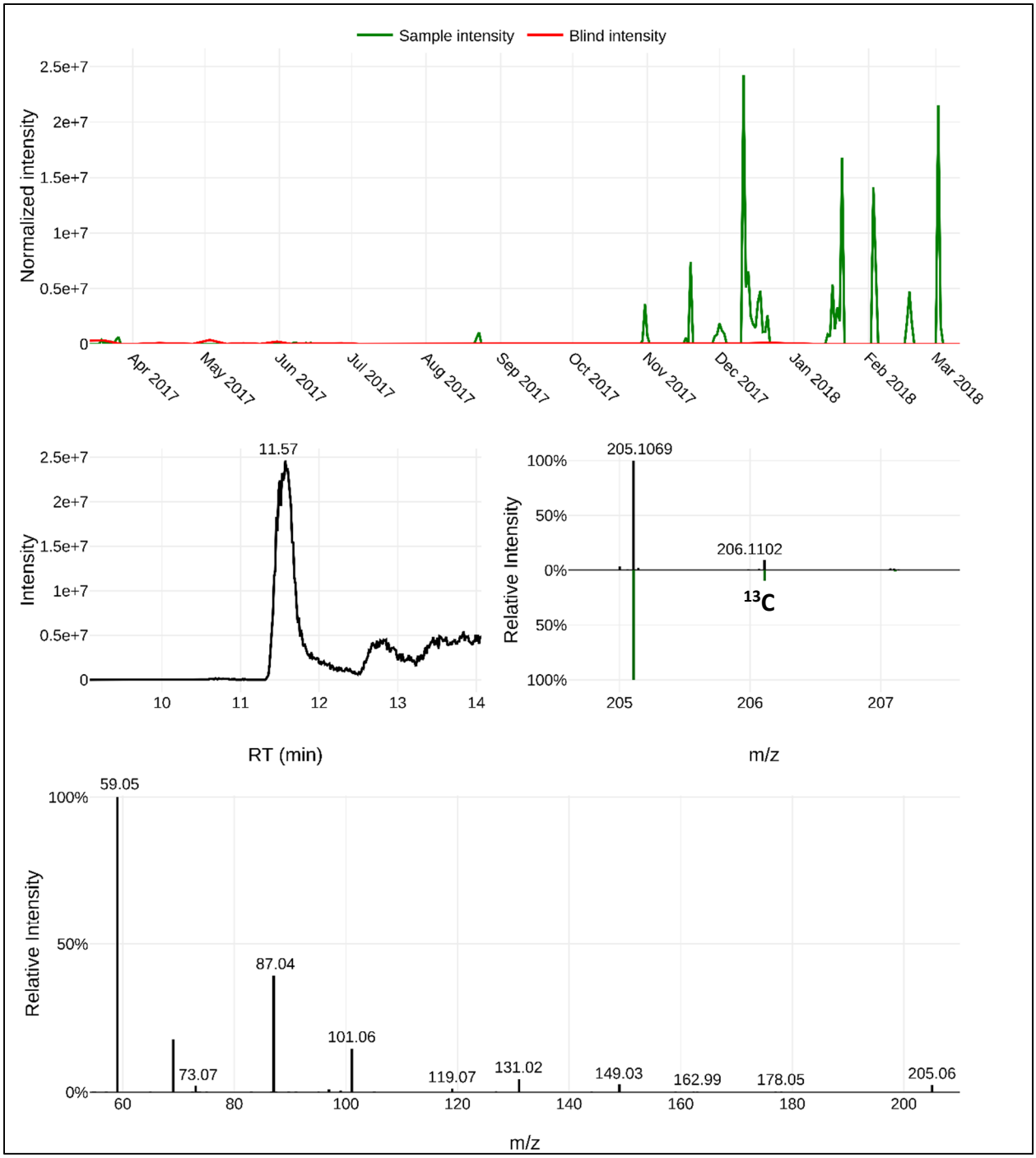



**Figure B65:** Identification of NT205b – level 4.

NT205b was identified to have an unequivocal molecular formula ($C_9H_{16}O_5$) which had a $RT_{avg}$ of 11.5 min. The proposed molecular formula assignment matches the MS spectra ($m/z_{avg}$ = 205.1071 for $[M+H]^+$, $\Delta m$ = 0.36 ppm), the theoretical abundance (9%) of the $^{13}C$ monoisotopic mass, and the $MS^2$ fragments.

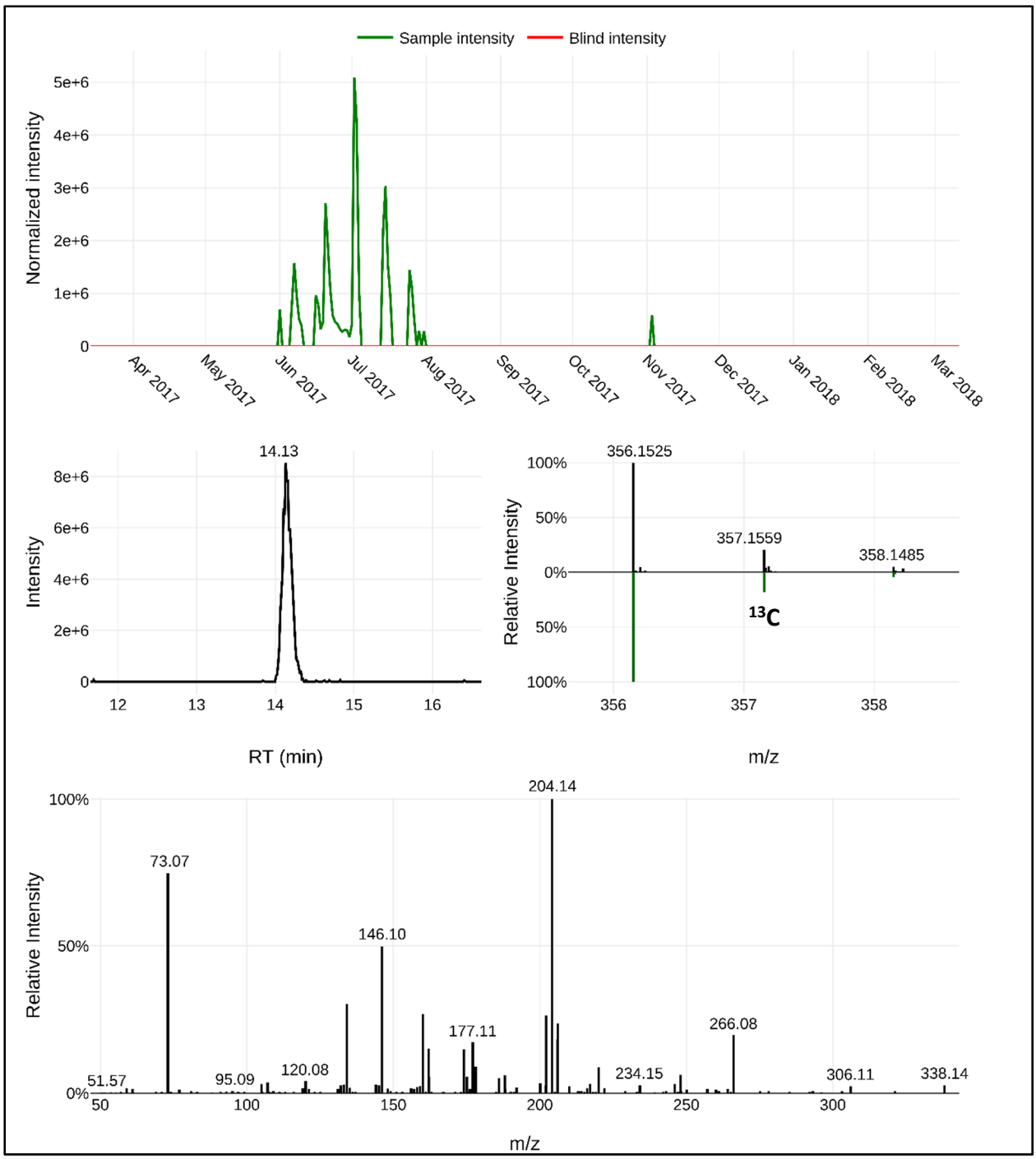



**Figure B66:** Identification of NT356 – level 4.

NT356 was identified to have an unequivocal molecular formula ($C_{17}H_{25}NO_5S$) which had a $RT_{avg}$ of 14.2 min. The proposed molecular formula assignment matches the MS spectra ($m/z_{avg}$ = 356.1527 for $[M+H]^+$, $\Delta m$ = 0.23 ppm), the theoretical abundance (17%) of the $^{13}C$ monoisotopic mass, the theoretical abundance (4.5%) of the $^{34}S$ monoisotopic mass, and the $MS^2$ fragments.

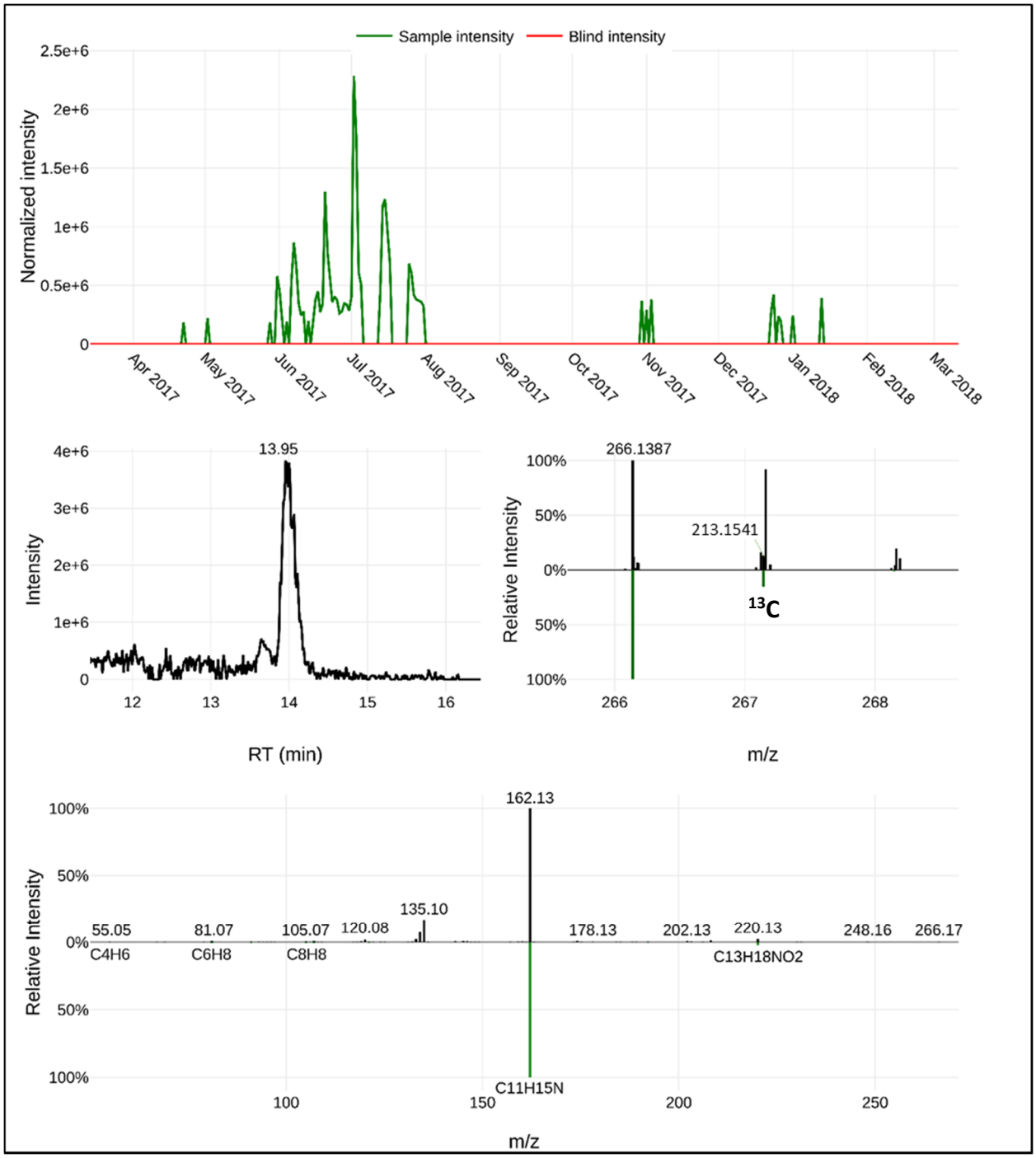



**Figure B67:** Identification of alachlor-OXA – level 1.

Alachlor-OXA ($C_{14}H_{19}NO_4$) was confirmed using an authentic reference standard which matched the $RT_{avg}$ of 14.0 min, the MS spectra ($m/z_{avg}$ = 266.1388 for $[M+H]^+$, $\Delta m$ = 0.31 ppm), the theoretical abundance (14%) of the $^{13}C$ monoisotopic mass, and the $MS^2$ fragments ($m/z$ = 162.13).

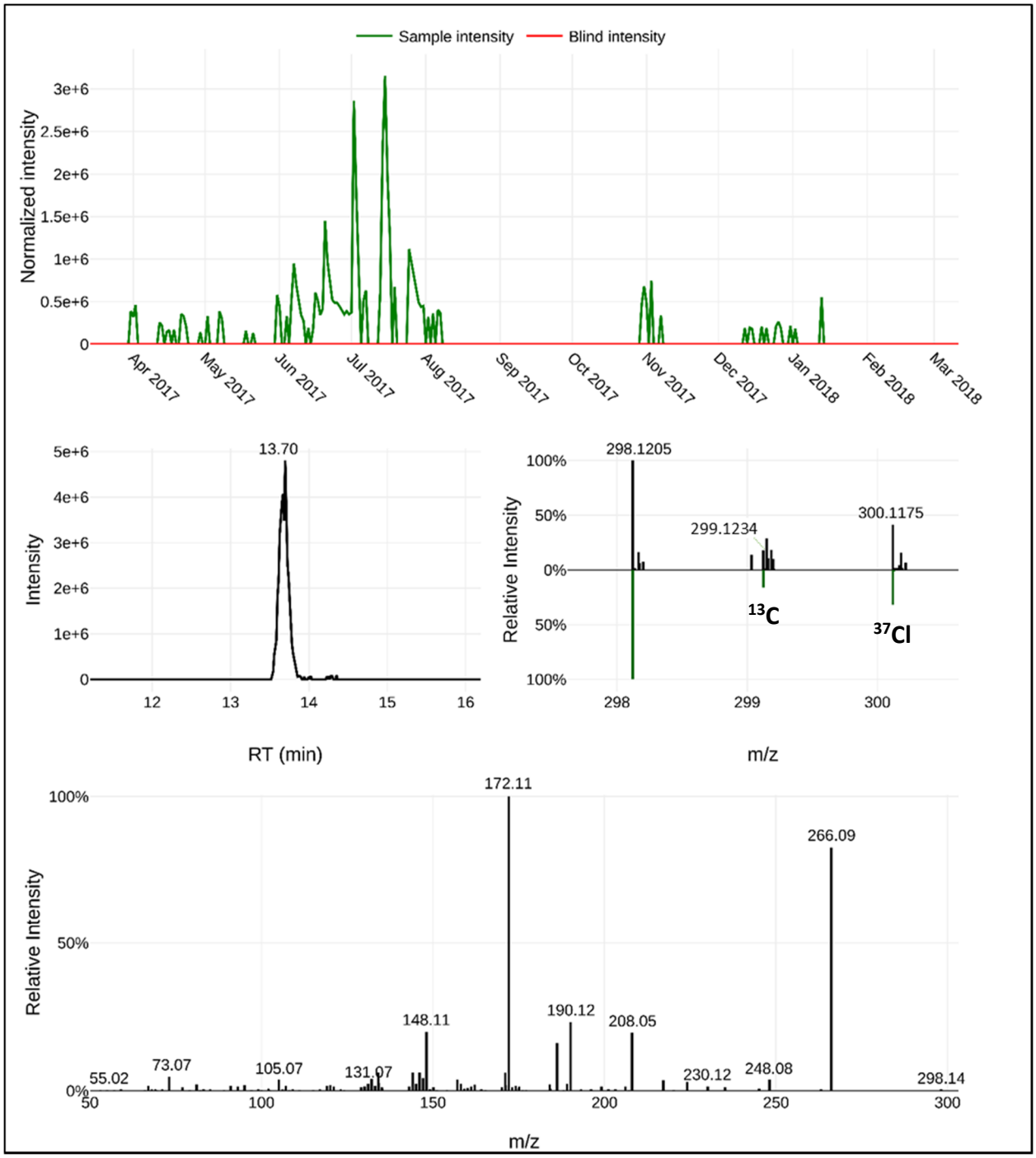



**Figure B68:** Identification of NT298 – level 4.

NT298 was identified to have an unequivocal molecular formula ($C_{15}H_{20}ClNO_3$) which had a $RT_{avg}$ of 13.7 min. The proposed molecular formula assignment matches the MS spectra ($m/z_{avg}$ = 298.1206 for $[M+H]^+$, $\Delta m$ = 0.36 ppm), the theoretical abundance (15%) of the $^{13}C$ monoisotopic mass, and the $MS^2$ fragments.

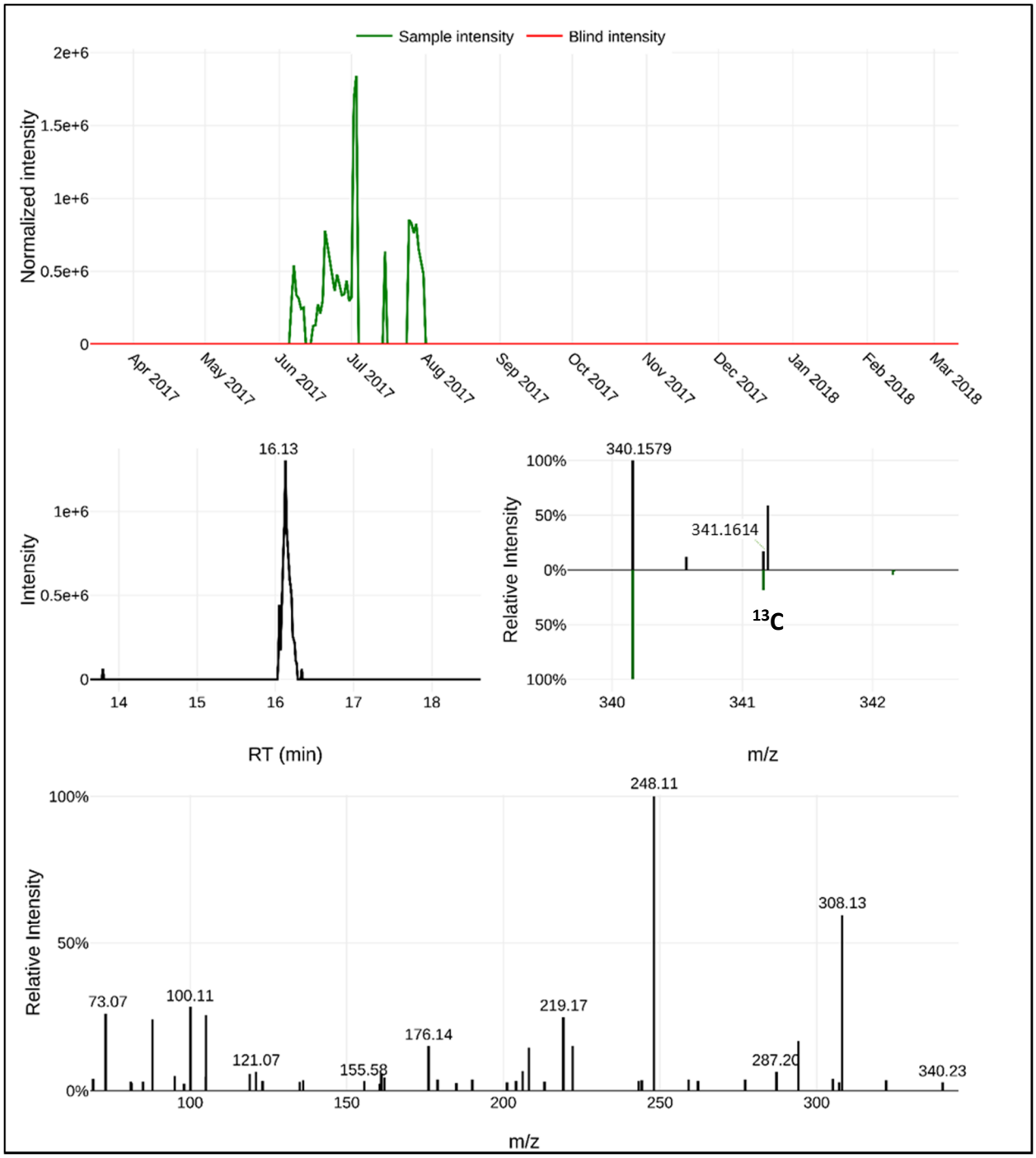



**Figure B69:** Identification of NT340 – level 4.

NT340 was identified to have an unequivocal molecular formula ($C_{17}H_{25}NO_4S$) which had a $RT_{avg}$ of 16.1 min. The proposed molecular formula assignment matches the MS spectra ($m/z_{avg}$ = 340.1578 for $[M+H]^+$, $\Delta m$ = 0.27 ppm), the theoretical abundance (17%) of the $^{13}C$ monoisotopic mass, and the $MS^2$ fragments.

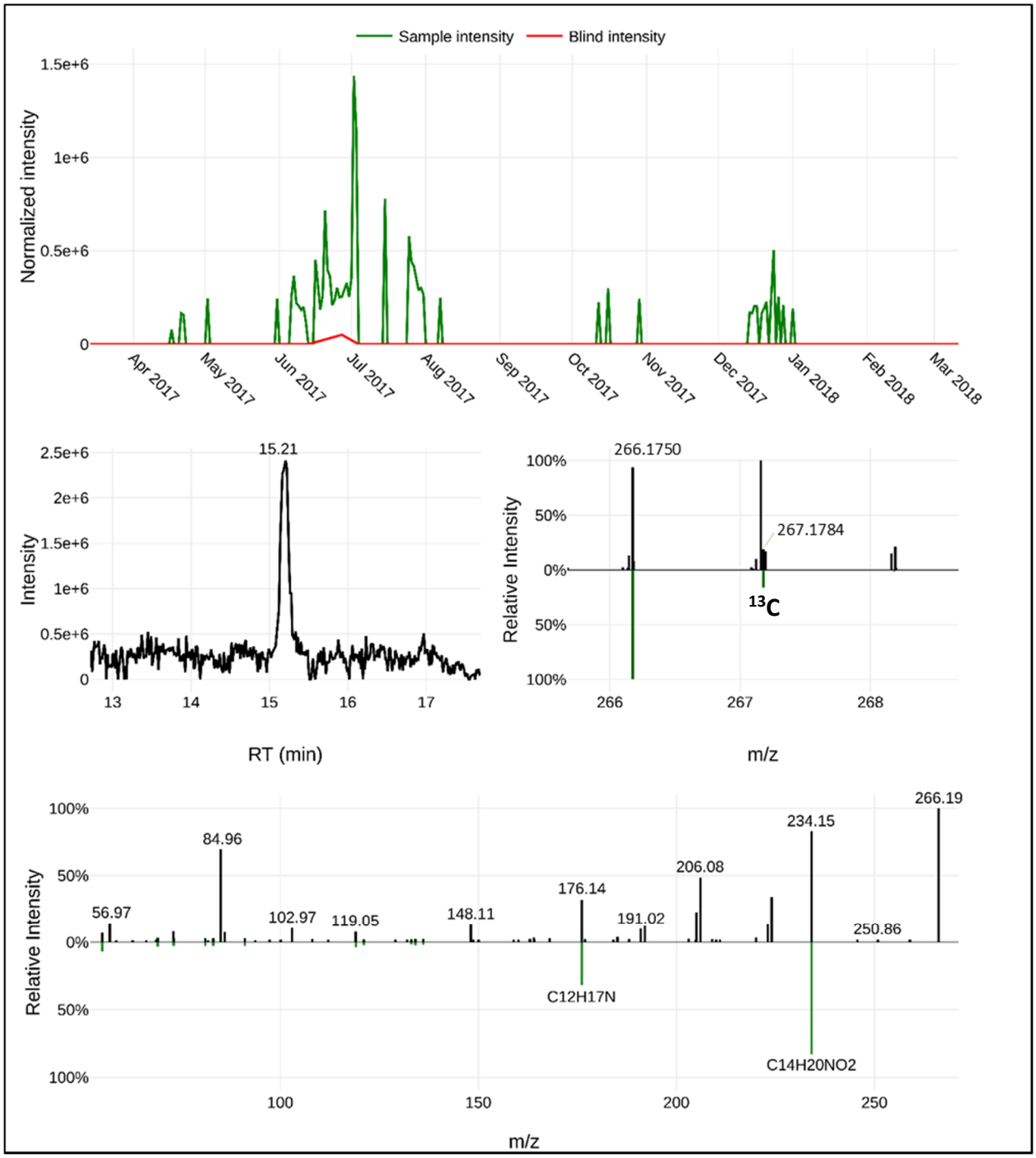



**Figure B70:** Identification of NT266 – level 2P.

NT266 likely represents metolachlor-2-hydroxy ($C_{15}H_{23}NO_3$) which had a $RT_{avg}$ of 15.2 min. The proposed MS assignment matches the MS spectra ($m/z_{avg}$ = 266.1752 for $[M+H]^+$, $\Delta m$ = 0.50 ppm), the theoretical abundance (15%) of the $^{13}C$ monoisotopic mass, and the $MS^2$ fragments generated with Metfrag.

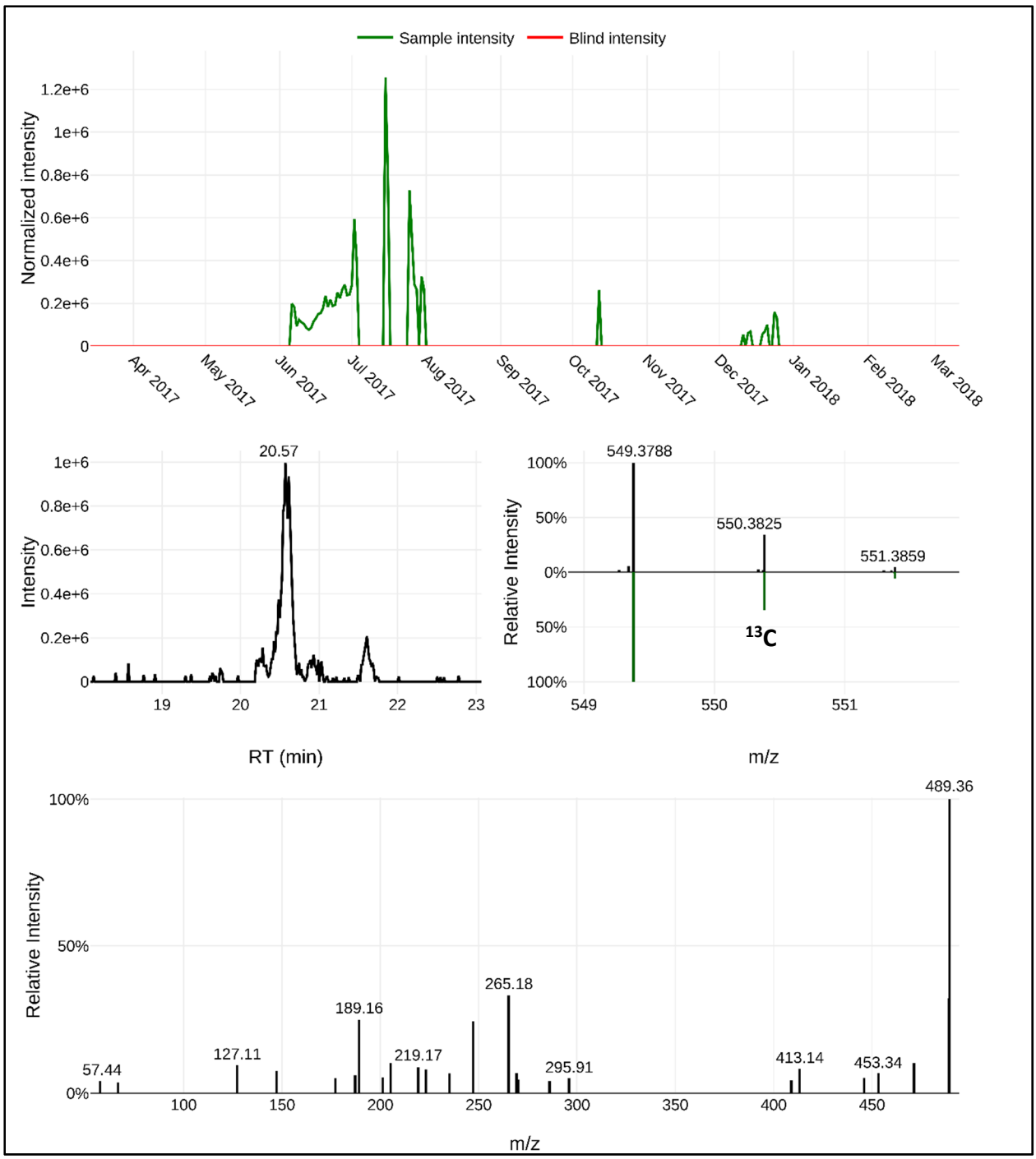



**Figure B71:** Identification of NT549 – level 4.

NT549 was identified to have an unequivocal molecular formula ($C_{32}H_{52}O_7$) which had a $RT_{avg}$ of 20.6 min. The proposed molecular formula assignment matches the MS spectra ($m/z_{avg}$ = 549.3782 for $[M+H]^+$, $\Delta m$ = -0.72 ppm), the theoretical abundance (32%) of the $^{13}C$ monoisotopic mass, and the $MS^2$ fragments.

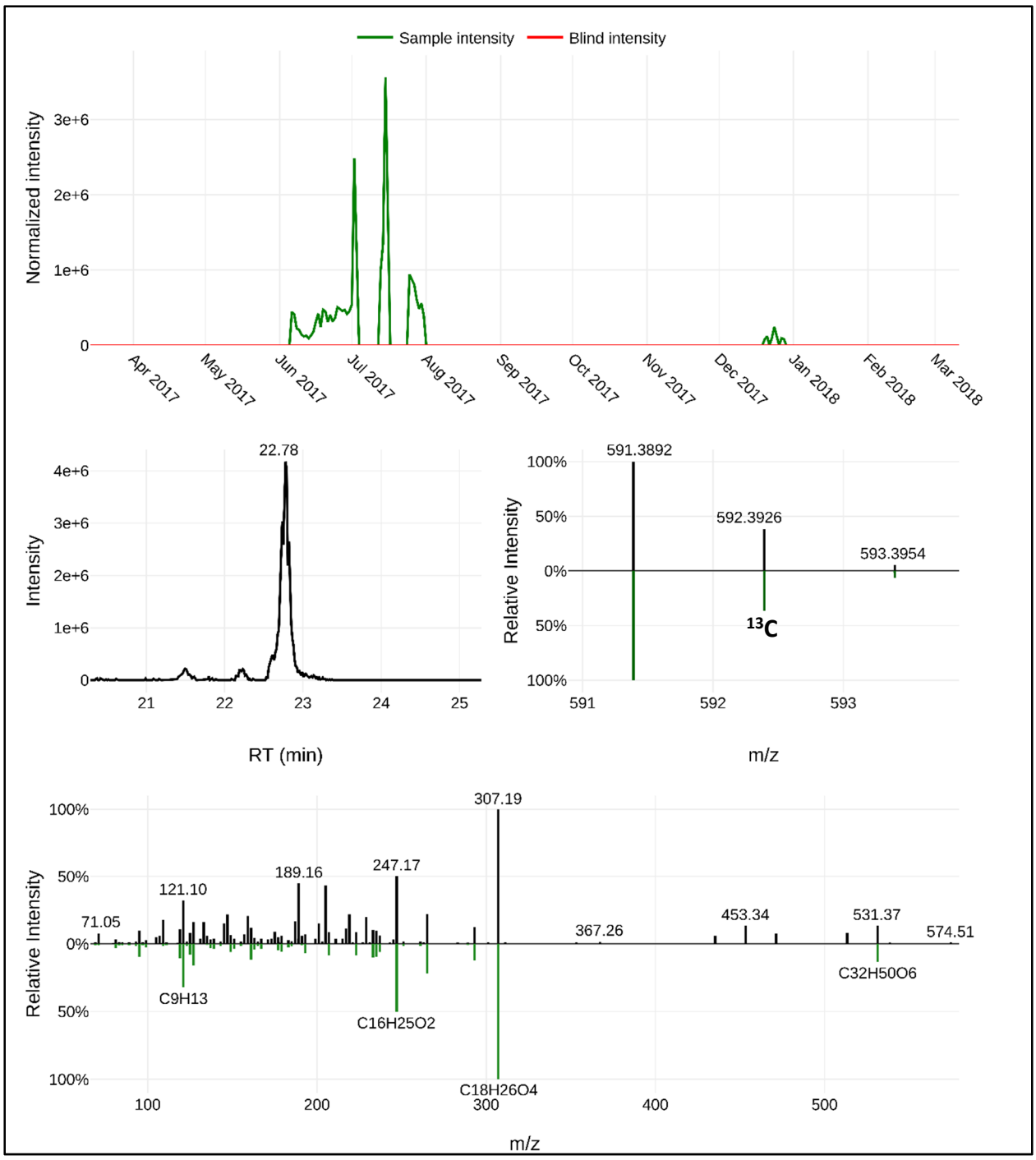



**Figure B72:** Identification of NT591 – level 4.

NT591 was identified to have an unequivocal molecular formula ($C_{34}H_{54}O_8$) which had a $RT_{avg}$ of 22.8 min. The proposed molecular formula assignment matches the MS spectra ($m/z_{avg}$ = 591.3886 for $[M+H]^+$, $\Delta m$ = -0.90 ppm), the theoretical abundance (34%) of the $^{13}C$ monoisotopic mass, and the $MS^2$ fragments.

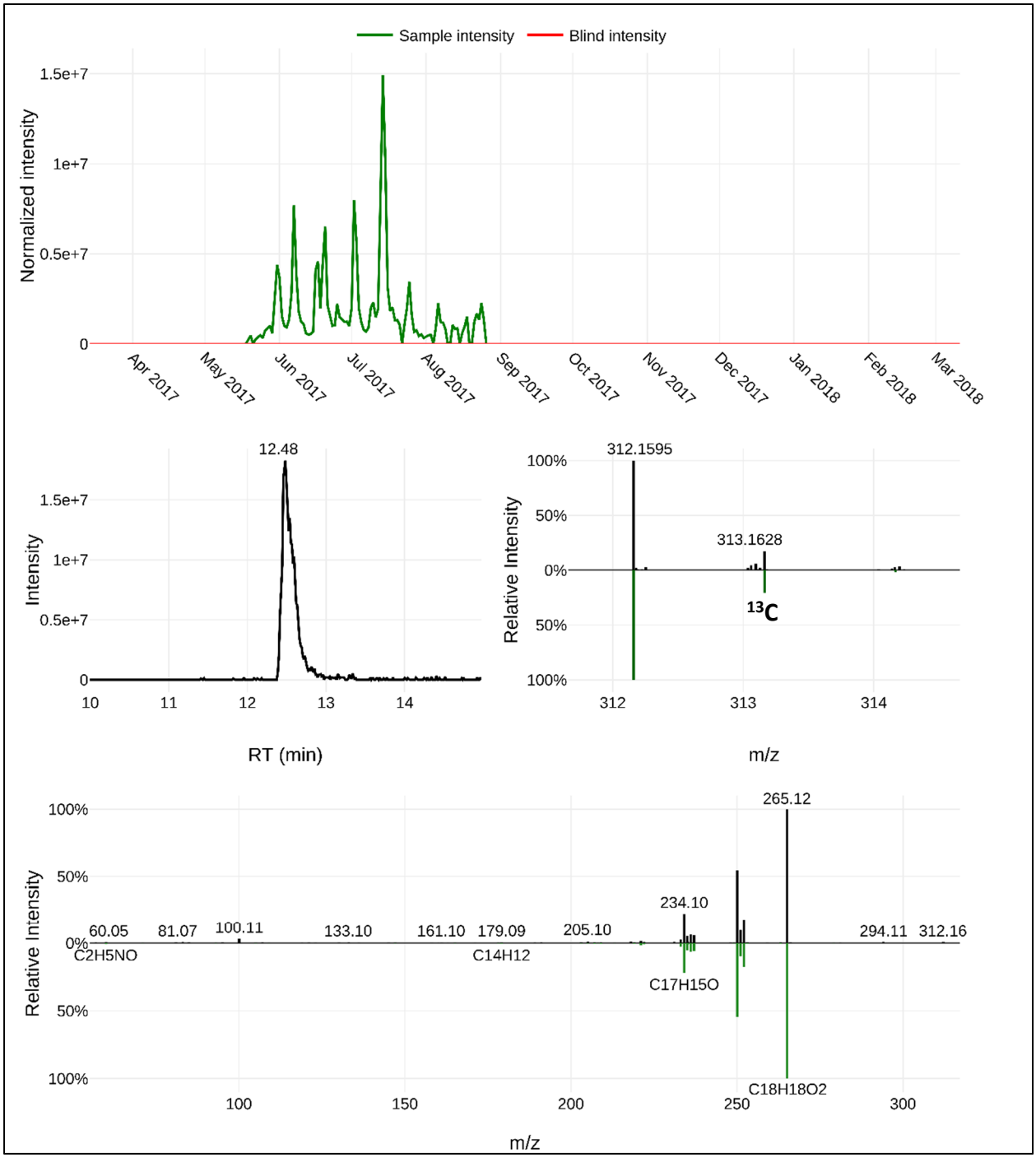



**Figure B73:** Identification of NT312 – level 2P.

NT312 likely represents nuciferine N-oxide ($C_{19}H_{21}NO_3$) which had a $RT_{avg}$ of 12.5 min. The proposed MS assignment matches the MS spectra ($m/z_{avg}$ = 312.1594 for $[M+H]^+$, $\Delta m$ = 0.05 ppm), the theoretical abundance (19%) of the $^{13}C$ monoisotopic mass, and the *in silico* $MS^2$ fragments generated with Metfrag.

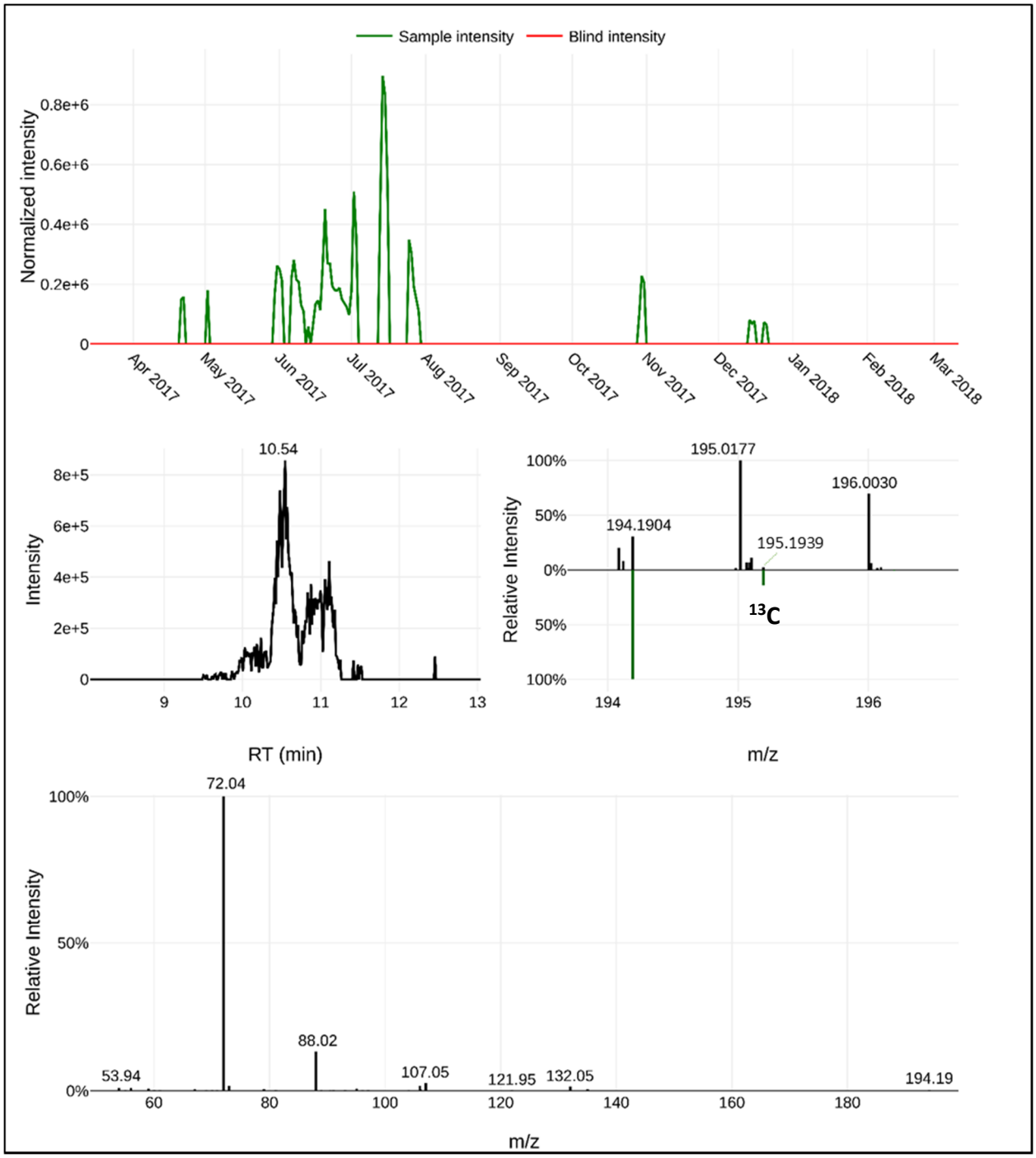



**Figure B74:** Identification of NT194 – level 4.

NT194 was identified to have an unequivocal molecular formula ($C_{13}H_{23}N$) which had a $RT_{avg}$ of 10.5 min. The proposed molecular formula assignment matches the MS spectra ($m/z_{avg}$ = 194.1905 for $[M+H]^+$, $\Delta m$ = 0.75 ppm), the theoretical abundance (13%) of the $^{13}C$ monoisotopic mass, and the $MS^2$ fragments.

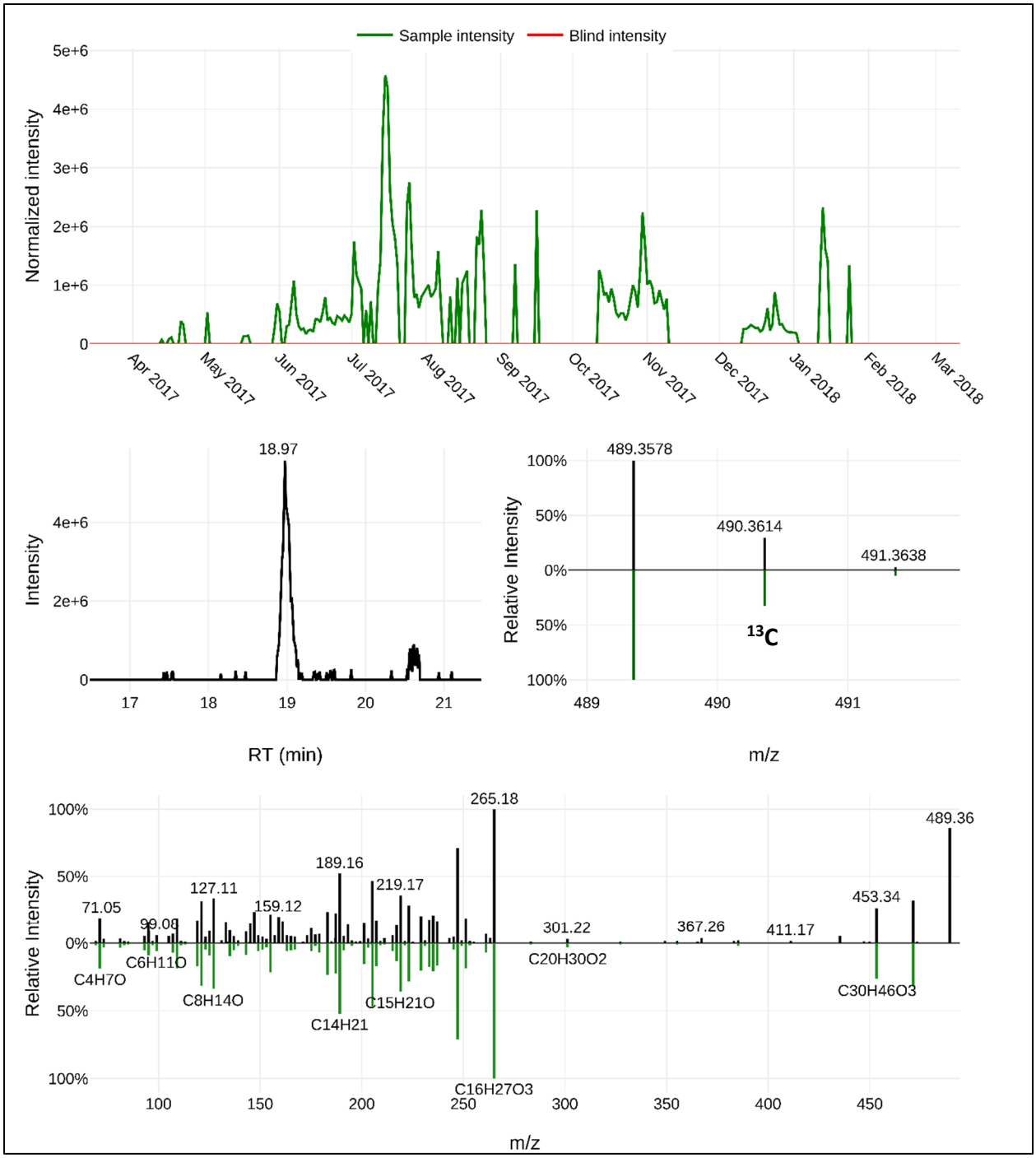



**Figure B75:** Identification of NT489 – level 2P.

NT489 likely represents esculentic acid ($C_{30}H_{48}O_5$) which had a $RT_{avg}$ of 19.0 min. The proposed MS assignment matches the MS spectra ($m/z_{avg}$ = 489.3573 for $[M+H]^+$, $\Delta m$ = -0.37 ppm), the theoretical abundance (30%) of the $^{13}C$ monoisotopic mass, and the *in silico* $MS^2$ fragments generated with Metfrag.

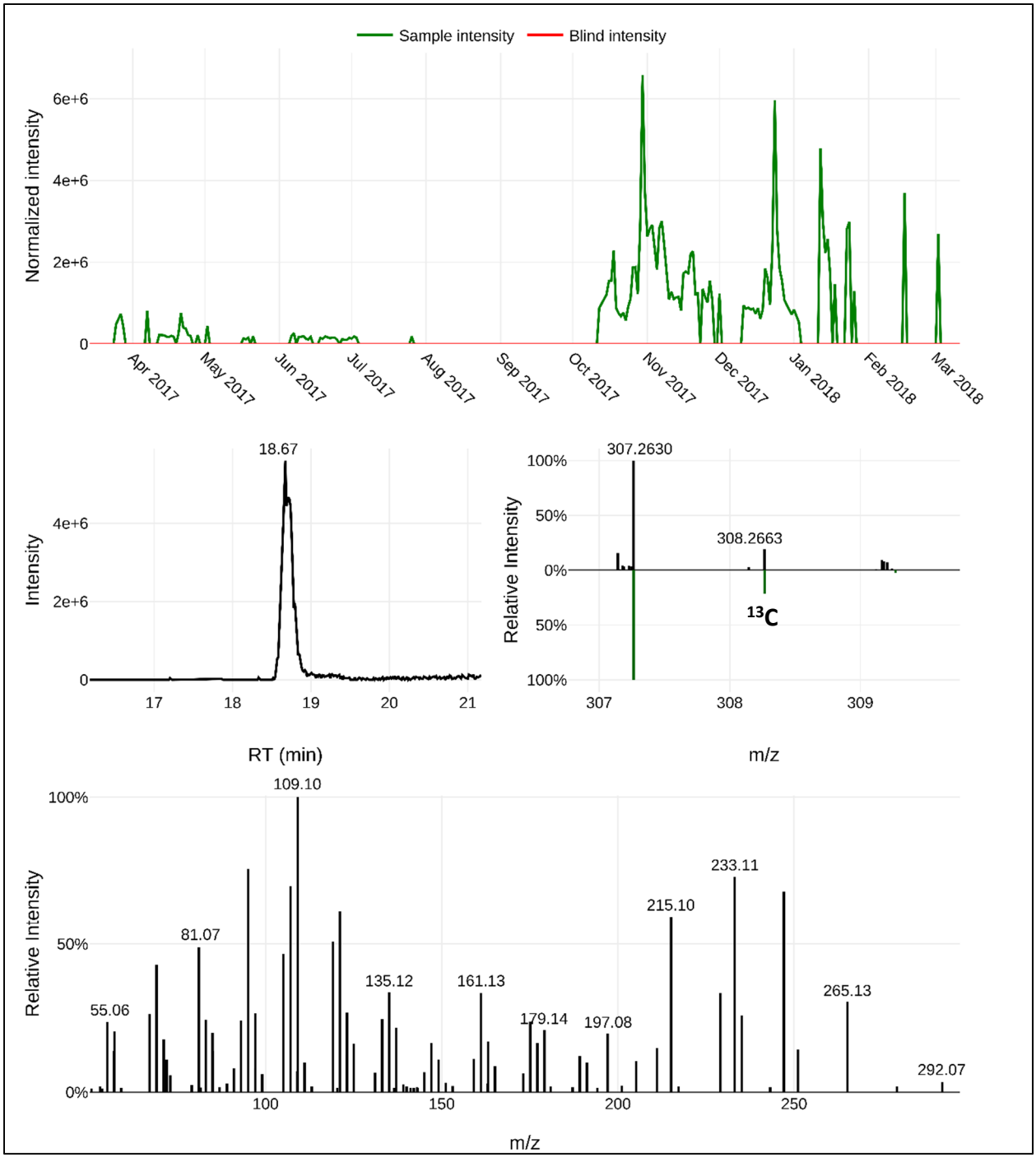



**Figure B76:** Identification of NT307 – level 4.

NT307 was identified to have an unequivocal molecular formula ($C_{20}H_{34}O_2$) which had a $RT_{avg}$ of 18.7 min. The proposed molecular formula assignment matches the MS spectra ($m/z_{avg}$ = 307.2631 for $[M+H]^+$, $\Delta m$ = -0.15 ppm), the theoretical abundance (20%) of the $^{13}C$ monoisotopic mass, and the $MS^2$ fragments.

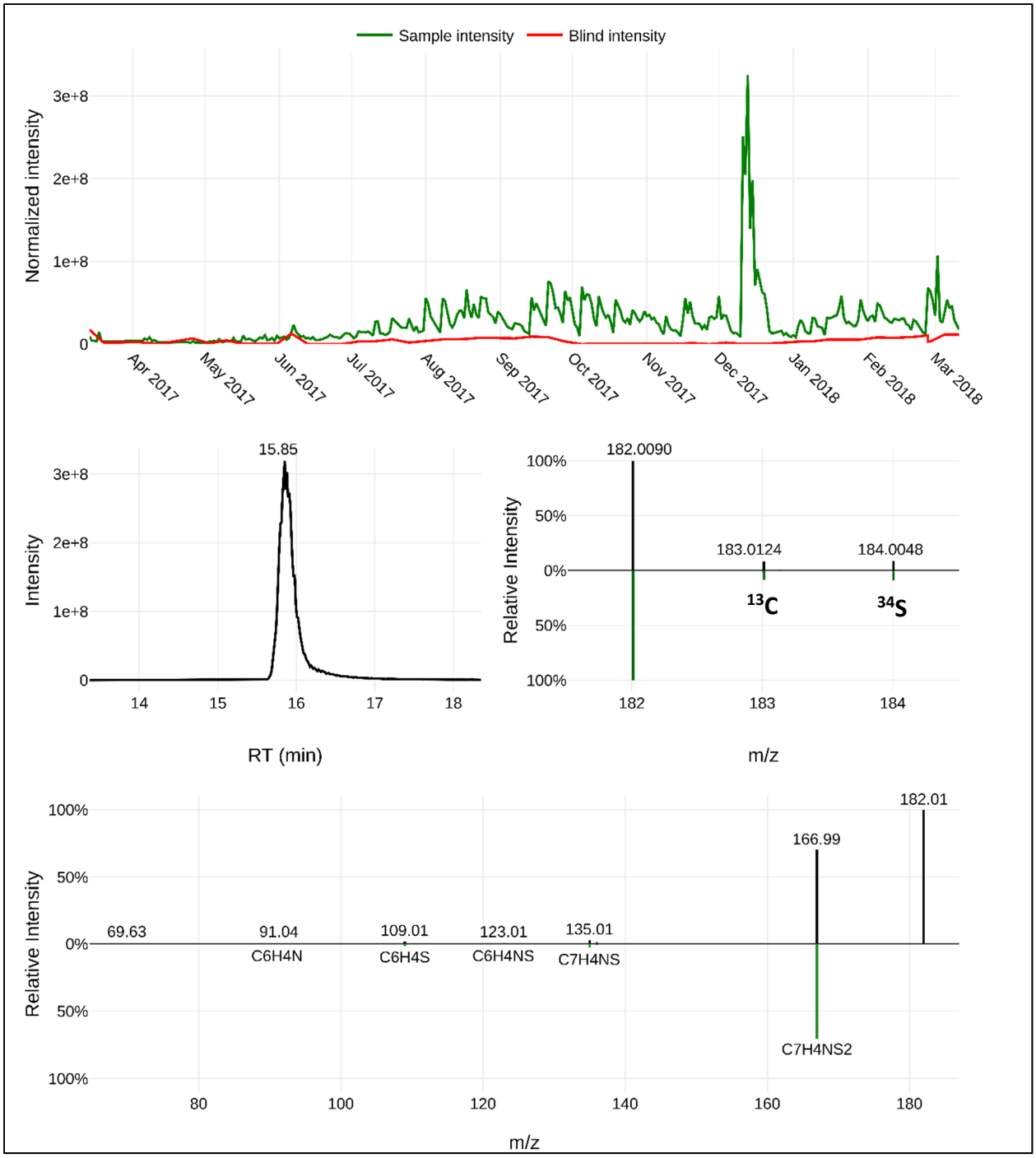



**Figure B77:** Identification of 2-methylthiobenzothiazole – level 1.

2-Methylthiobenzothiazole ($C_8H_7NS_2$) was confirmed using an authentic reference standard which matched the $RT_{avg}$ of 15.8 min, the MS spectra ($m/z_{avg}$ = 182.0092 for $[M+H]^+$, $\Delta m$ = -0.46 ppm), the theoretical abundance (8%) of the $^{13}C$ monoisotopic mass, the theoretical abundance (9%) of the $^{34}S$ monoisotopic mass, and the $MS^2$ fragments ($m/z$ = 109.01, 135.01, and 166.99).

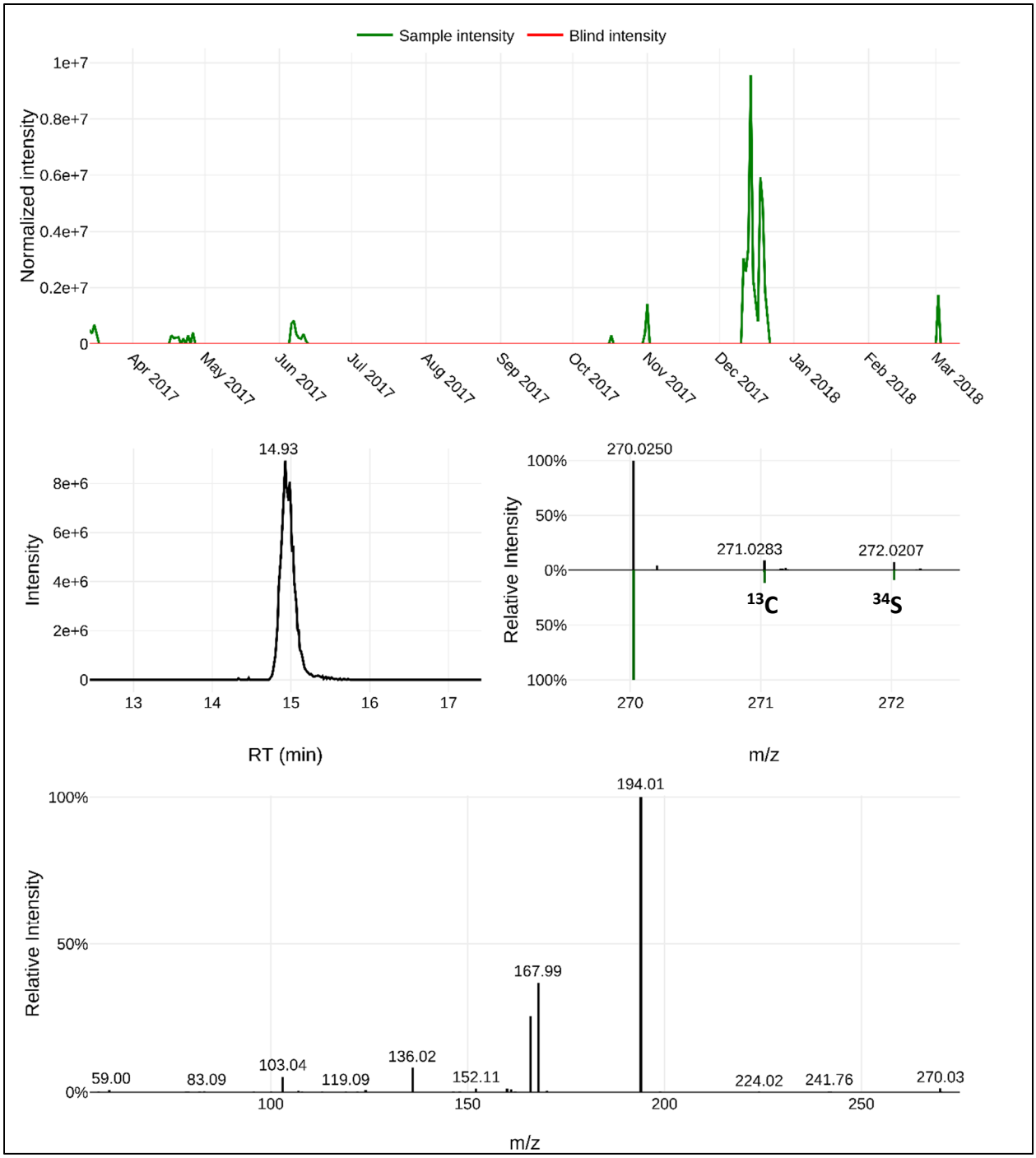



**Figure B78:** Identification of NT270 – level 3.

NT270 was identified to have an unequivocal molecular formula ($C_{11}H_{11}NO_3S_2$) which had a $RT_{avg}$ of 14.9 min. The proposed molecular formula assignment matches the MS spectra ($m/z_{avg}$ = 270.0253 for $[M+H]^+$, $\Delta m$ = -0.06 ppm), the theoretical abundance (11%) of the $^{13}C$ monoisotopic mass, the theoretical abundance (9%) of the $^{34}S$ monoisotopic mass, and the $MS^2$ fragments. The $MS^2$ fragments ($m/z$ = 136.02 and 167.99) suggest a 2-mercaptobenzothiazole sub-structure.

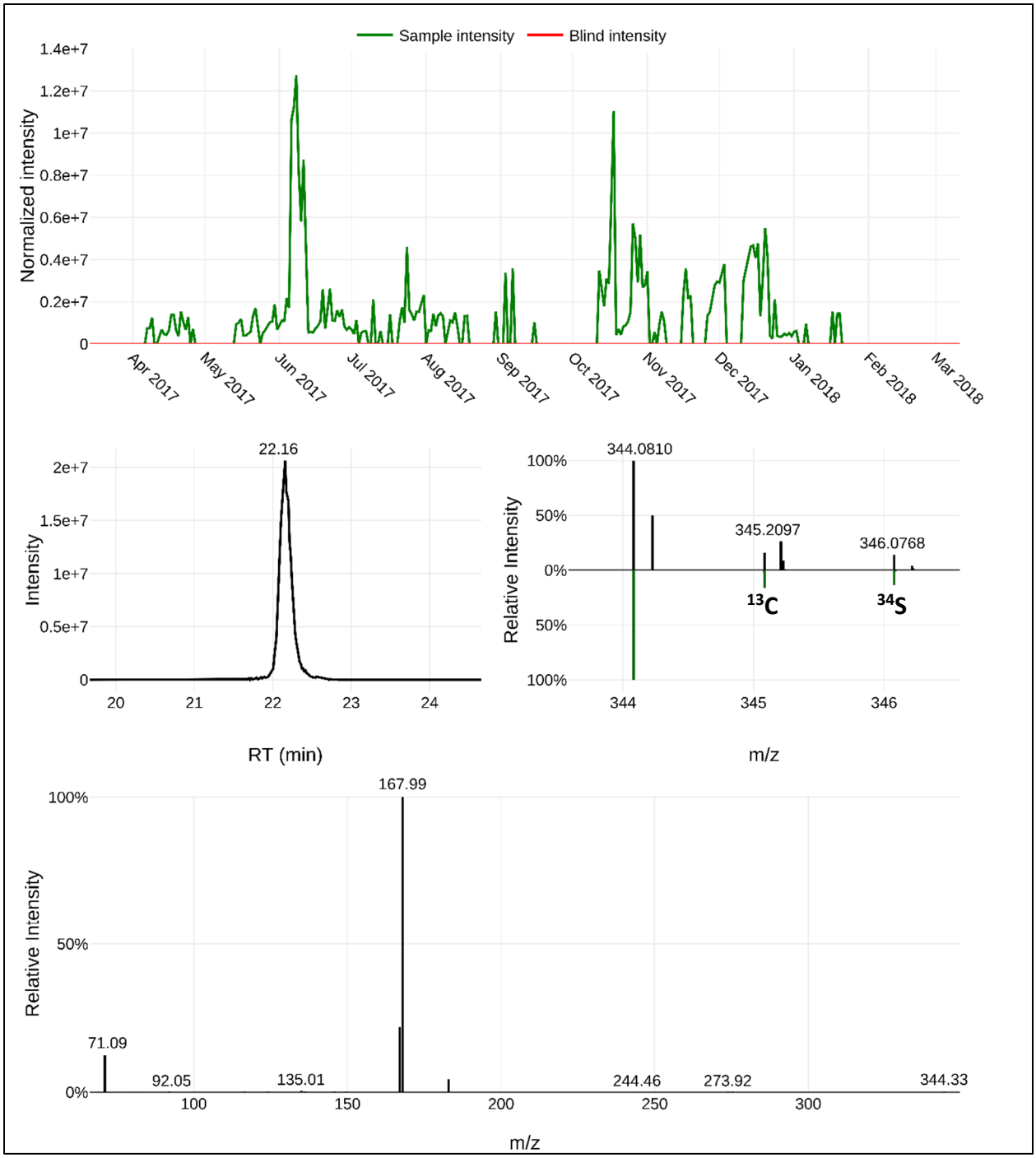



**Figure B79:** Identification of NT344 – level 3.

NT344 was identified to have an unequivocal molecular formula ($C_{15}H_{21}NO_2S_3$) which had a $RT_{avg}$ of 22.2 min. The proposed molecular formula assignment matches the MS spectra ($m/z_{avg}$ = 344.0807 for $[M+H]^+$, $\Delta m$ = -0.08 ppm), the theoretical abundance (15%) of the $^{13}C$ monoisotopic mass, the theoretical abundance (13.5%) of the $^{34}S$ monoisotopic mass, and the $MS^2$ fragments. The $MS^2$ fragments ($m/z$ = 135.01, 166.99, and 167.99) suggest a 2-mercaptobenzothiazole sub-structure.

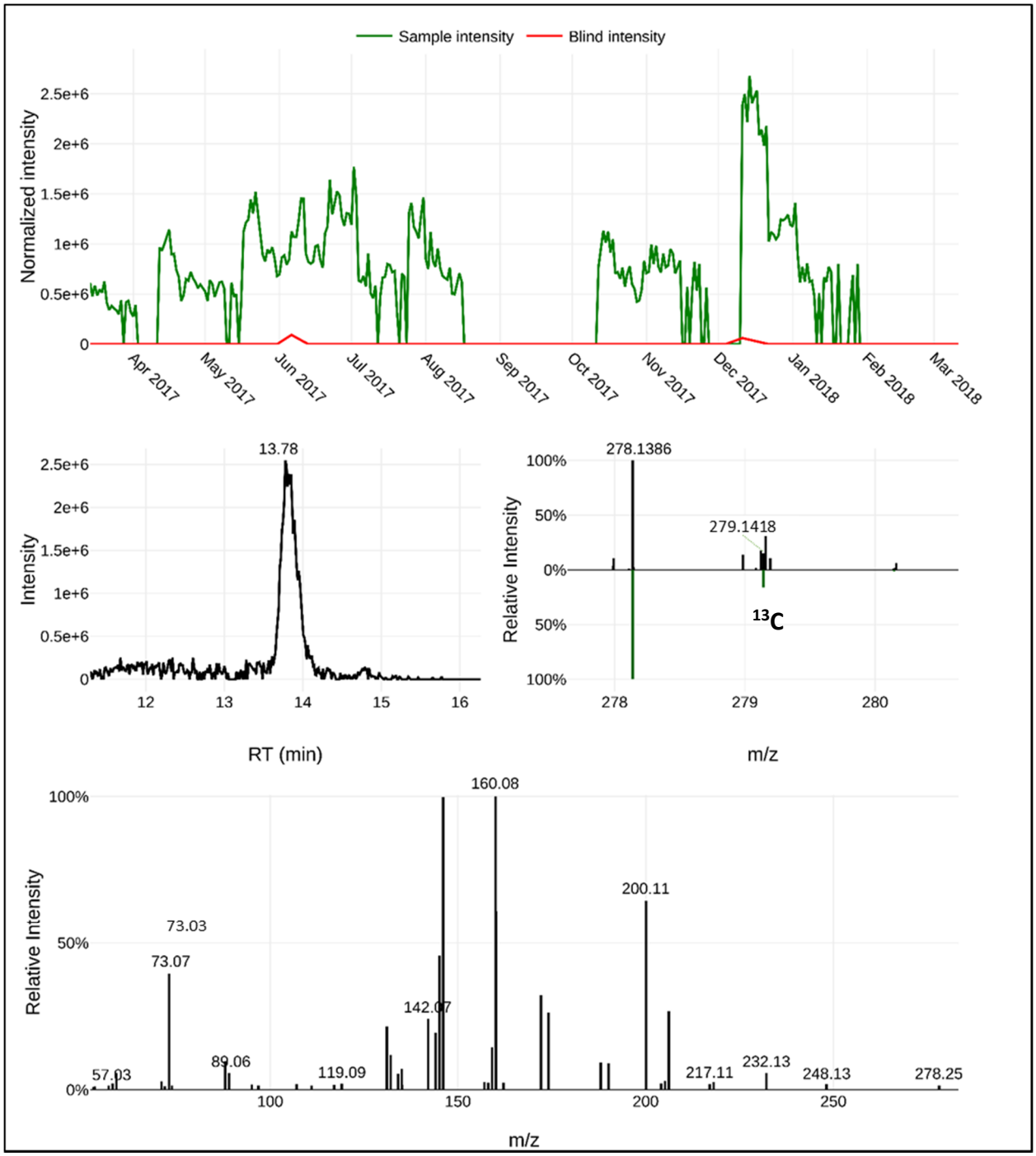



**Figure B80:** Identification of NT278b – level 4.

NT278b was identified to have an unequivocal molecular formula ($C_{15}H_{19}NO_4$) which had a $RT_{avg}$ of 13.7 min. The proposed molecular formula assignment matches the MS spectra ($m/z_{avg}$ = 278.1388 for $[M+H]^+$, $\Delta m$ = 0.54 ppm), the theoretical abundance (15%) of the $^{13}C$ monoisotopic mass, and the $MS^2$ fragments.

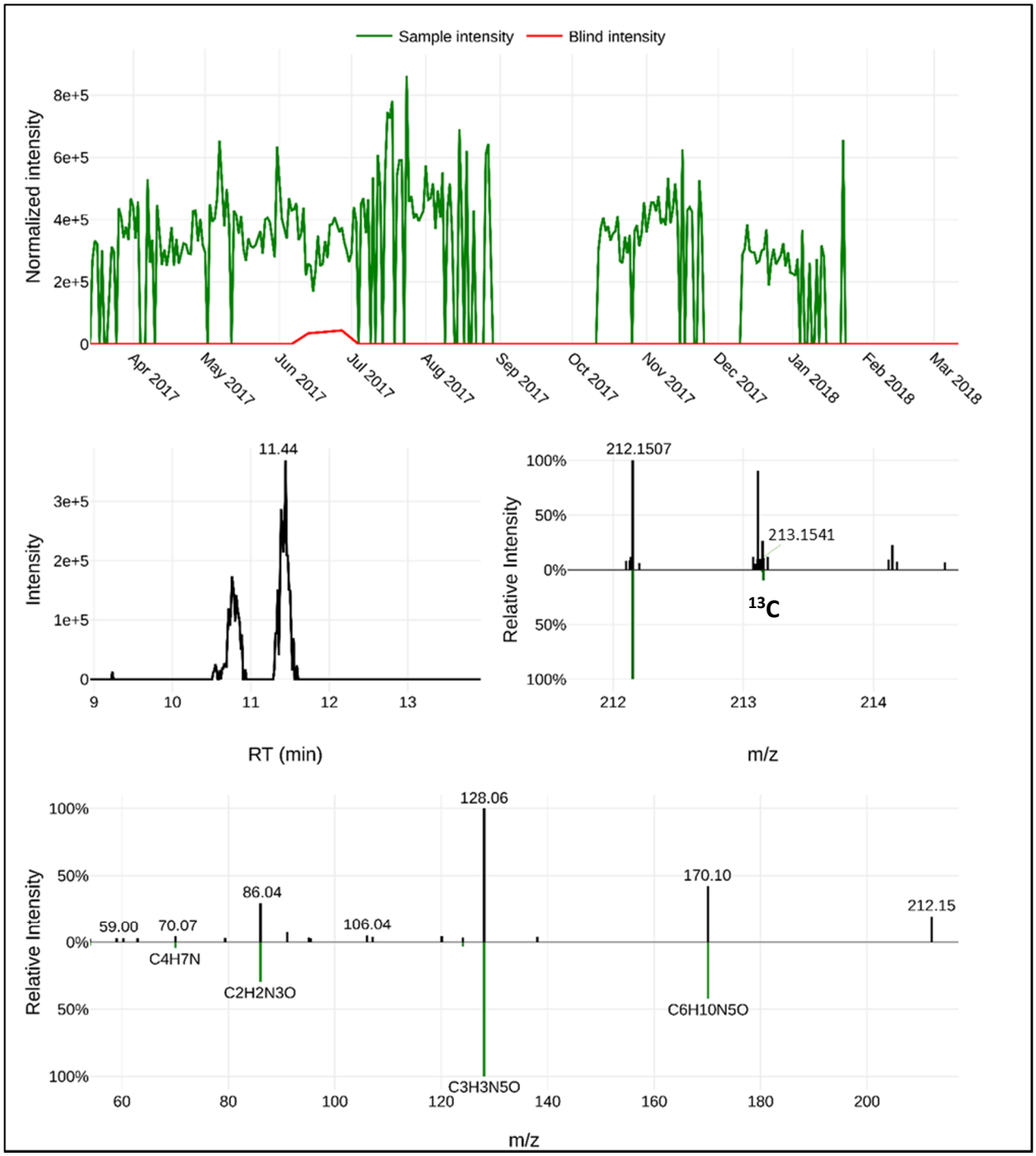



**Figure B81:** Identification of propazine-2-hydroxy – level 1L.

Propazine-2-hydroxy ($C_9H_{17}N_5O$) was using a library database which matched the $RT_{avg}$ of 11.4 min, the MS spectra ($m/z_{avg}$ = 212.1506 for $[M+H]^+$, Δm = 0.24 ppm), the theoretical abundance (9%) of the $^{13}C$ monoisotopic mass, and the $MS^2$ fragments ($m/z$ = 86.04, 128.06, and 170.10).

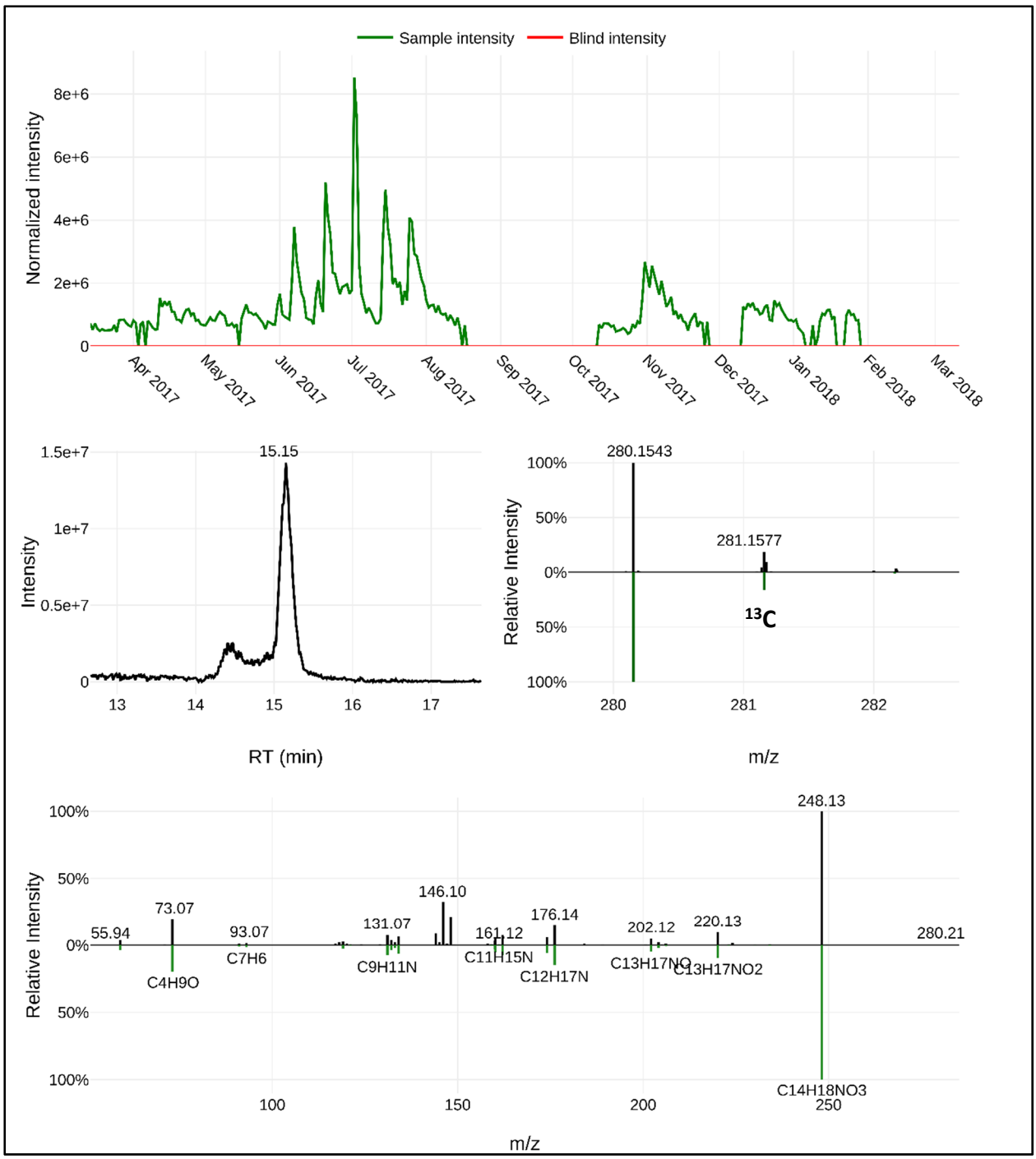



**Figure B82:** Identification of metolachlor-OXA – level 1.

Metolachlor-OXA ($C_{15}H_{21}NO_4$) was confirmed using an authentic reference standard which matched the $RT_{avg}$ of 15.2 min, the MS spectra ($m/z_{avg}$ = 280.1544 for $[M+H]^+$, $\Delta m$ = 0.21 ppm), the theoretical abundance (15%) of the $^{13}C$ monoisotopic mass, and the $MS^2$ fragments ($m/z$ = 73.07, 146.10, and 248.13).

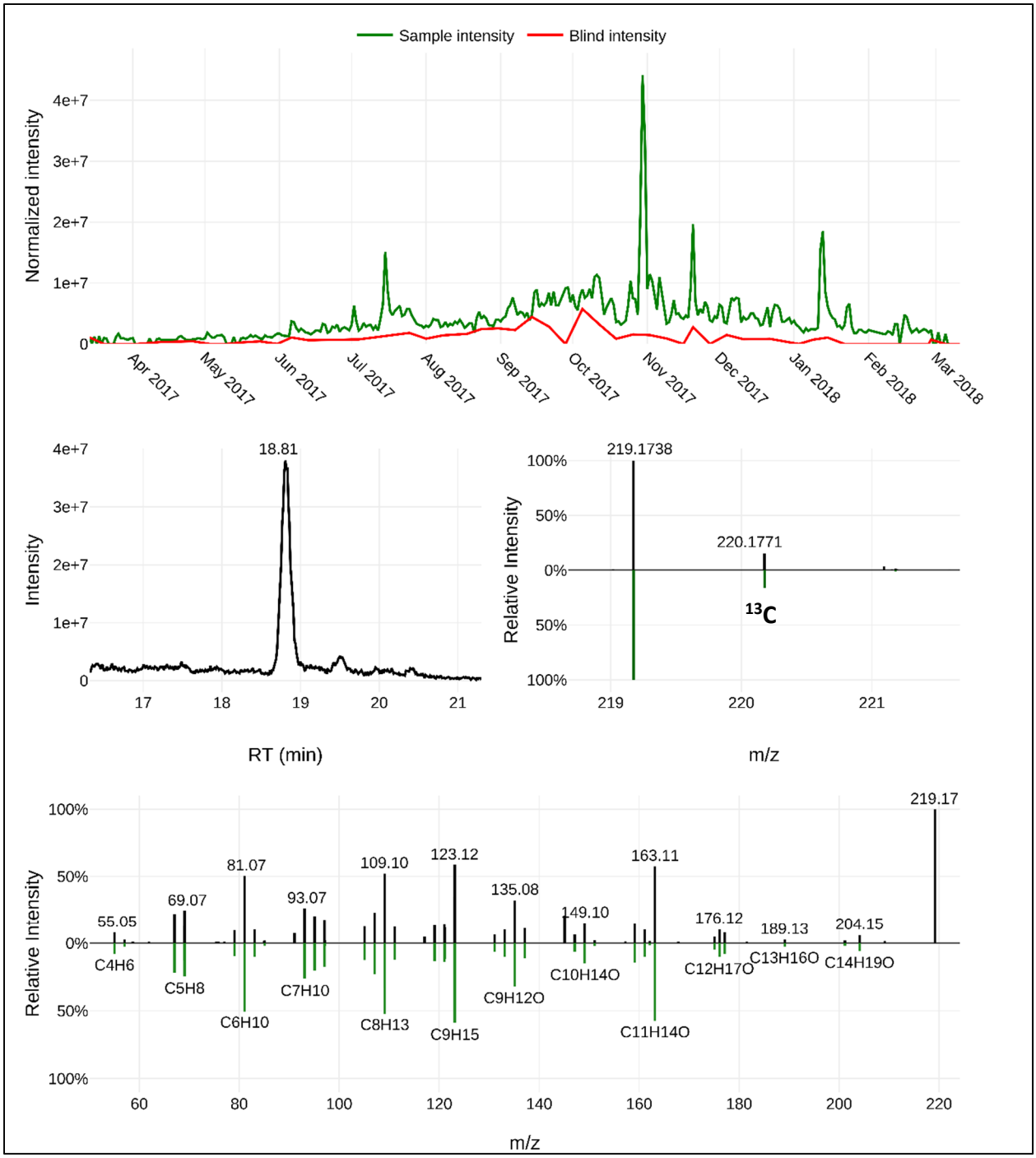



**Figure B83:** Identification of NT219 – level 2L.

NT219 likely represents nootkatone ($C_{15}H_{22}O$) which had a $RT_{avg}$ of 18.6 min. The proposed MS assignment matches the MS spectra ($m/z_{avg}$ = 219.1743 for $[M+H]^+$, Δm = -0.02 ppm), the theoretical abundance (15%) of the $^{13}C$ monoisotopic mass, and the $MS^2$ fragments on mzcloud ($m/z$ = 81.07, 123.12, and 163.11).

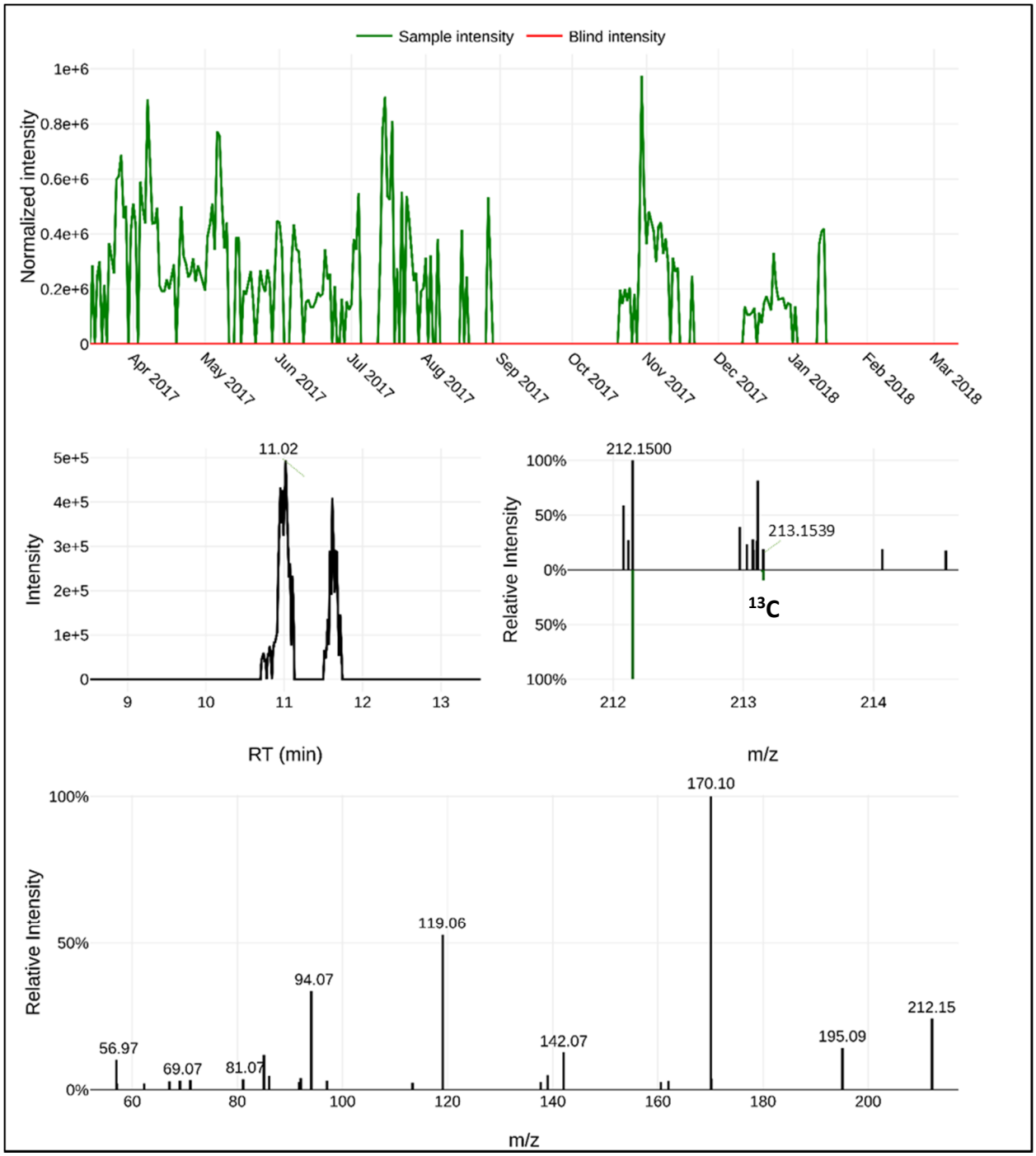



**Figure B84:** Identification of NT212 – level 3.

NT212 was identified to have an unequivocal molecular formula ($C_9H_{17}N_5O$) which had a $RT_{avg}$ of 10.7 min. The proposed molecular formula assignment matches the MS spectra ($m/z_{avg}$ = 212.1506 for $[M+H]^+$, $\Delta m$ = -0.07 ppm), the theoretical abundace (9%) of the $^{13}C$ monoisotopic mass, and the $MS^2$ fragments. The $MS^2$ fragments ($m/z$ = 86.03, 142.07, and 170.10) suggest a hydroxy-*s*-triazine sub-structure.

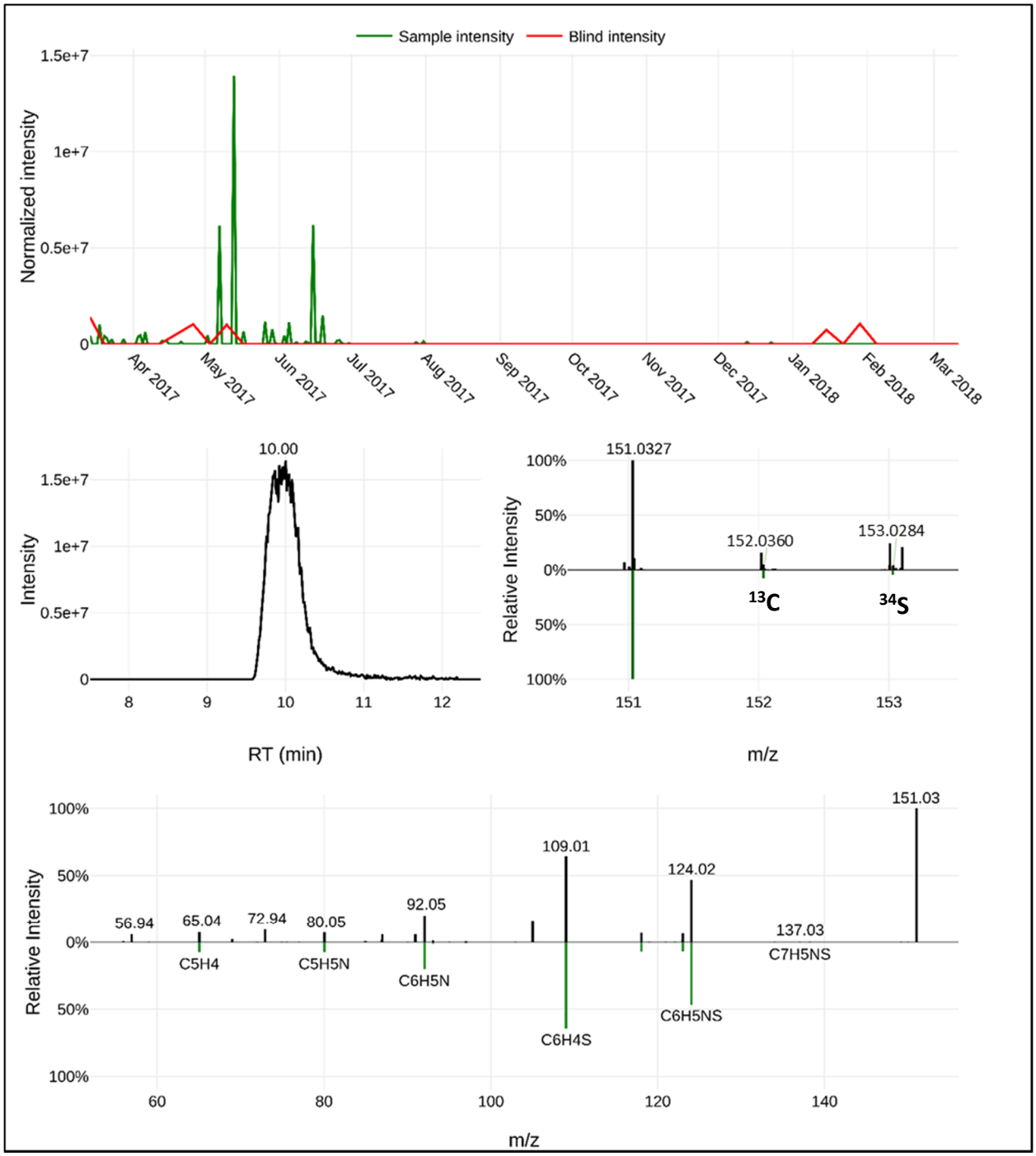



**Figure B85:** Identification of 2-aminobenzothiazole – level 1.

2-aminobenzothiazole ($C_7H_6N_2S$) was confirmed using an authentic reference standard which matched the $RT_{avg}$ of 10.1 min, the MS spectra ($m/z_{avg}$ = 151.0324 for $[M+H]^+$, $\Delta m$ = -0.47 ppm), the theoretical abundance (7%) of the $^{13}C$ monoisotopic mass, the theoretical abundance (4.5%) of the $^{34}S$ monoisotopic mass, and the $MS^2$ fragments ($m/z$ = 92.05, 109.01, and 124.02).

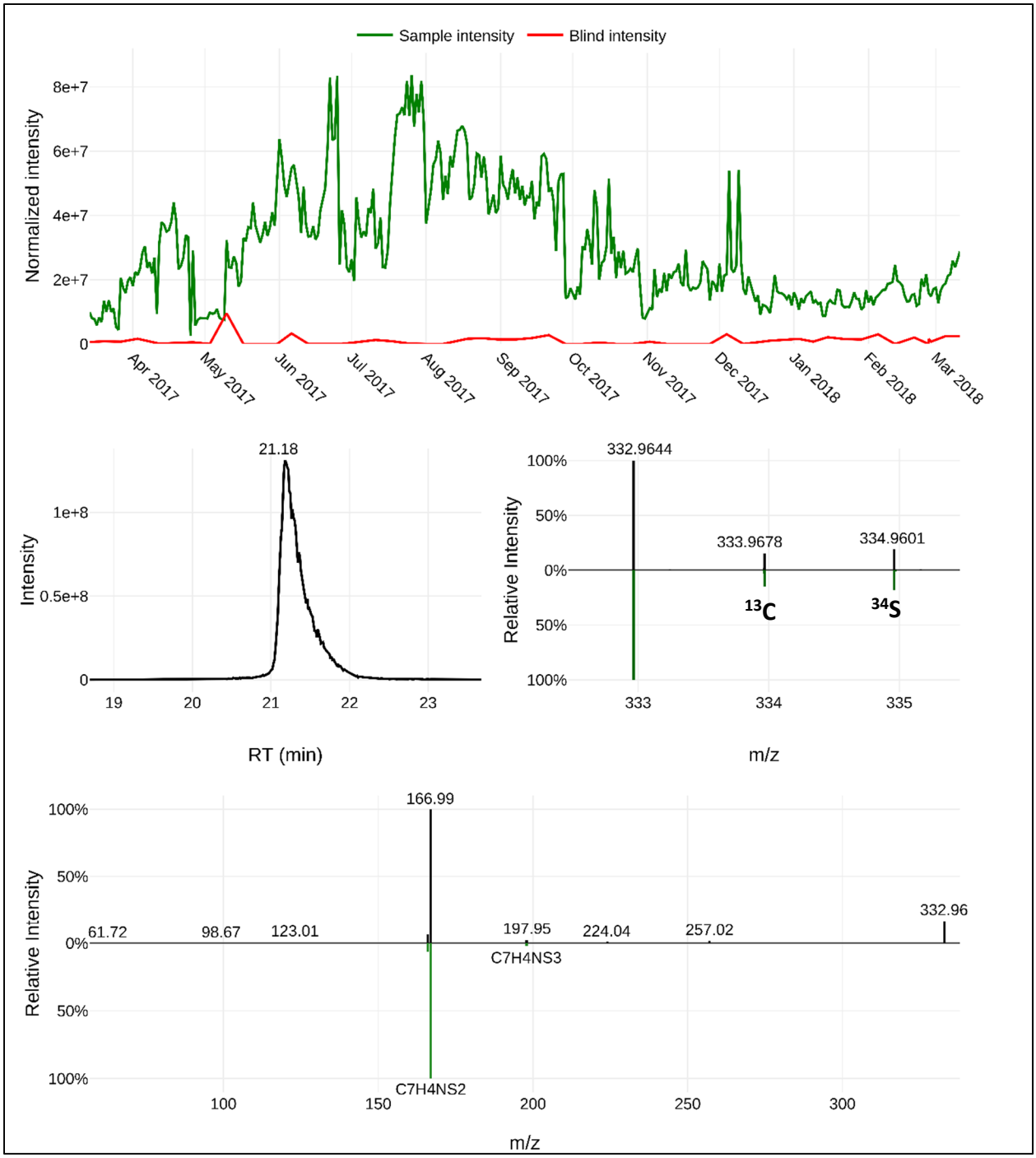



**Figure B86:** Identification of 2,2'-dithiobisbenzothiazole – level 1.

2,2'-Dithiobisbenzothiazole ($C_{14}H_8N_2S_4$) was confirmed using an authentic reference standard which matched the $RT_{avg}$ of 21.2 min, the MS spectra ($m/z_{avg}$ = 332.9642 for $[M+H]^+$, $\Delta m$ = -0.23 ppm), the theoretical abundance (14%) of the $^{13}C$ monoisotopic mass, the theoretical abundance (18%) of the $^{34}S$ monoisotopic mass, and the $MS^2$ fragments ($m/z$ = 165.98 and 166.99).

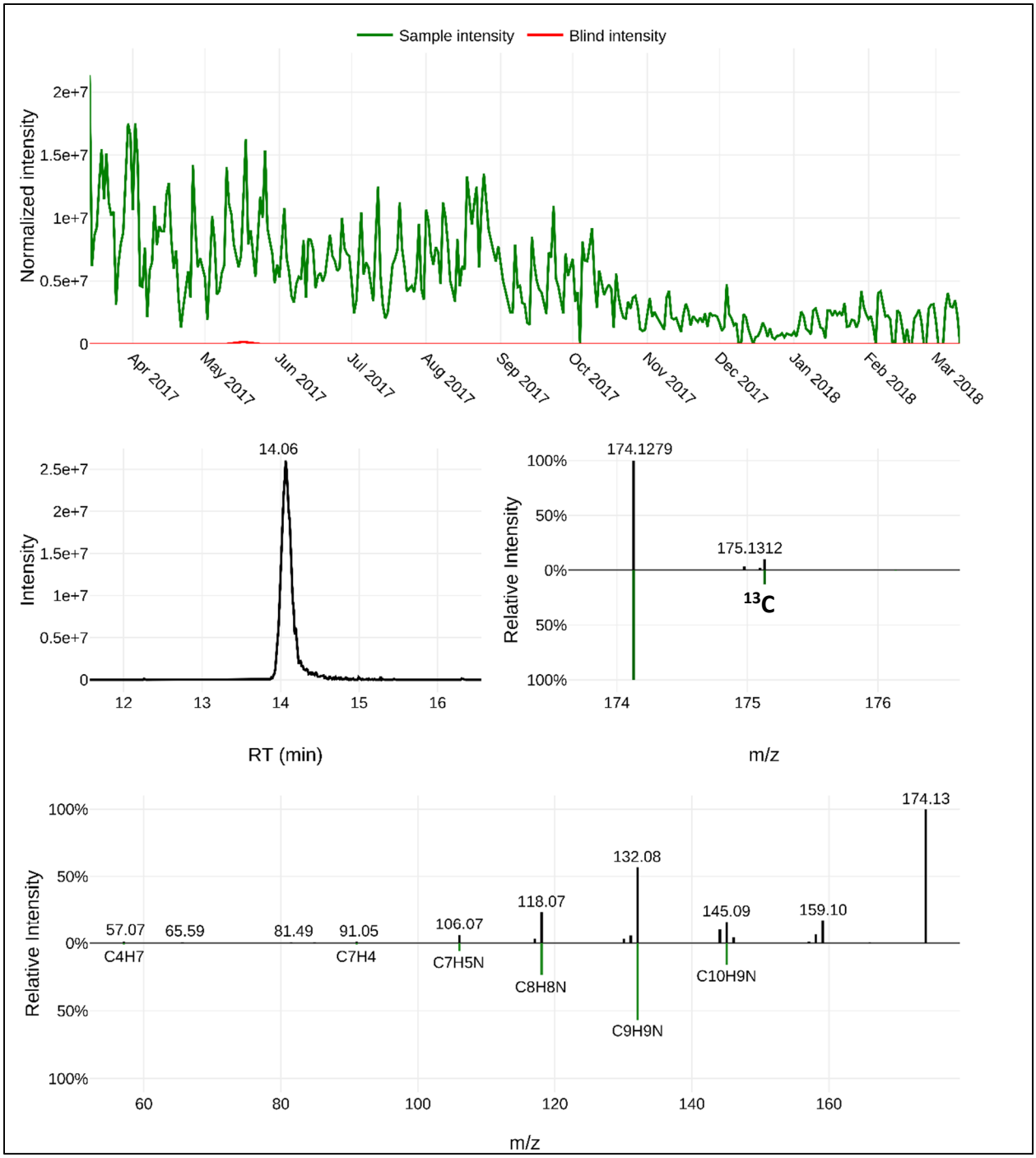



**Figure B87:** Identification of 1,2-dihydro-2,2,4-trimethylquinoline – level 1.

1,2-dihydro-2,2,4-trimethylquinoline ($C_{12}H_{15}N$) was confirmed using an authentic reference standard which matched the $RT_{avg}$ of 14.2 min, the MS spectra spectra ($m/z_{avg}$ = 174.1277 for $[M+H]^+$, $\Delta m$ = 0.07 ppm), the theoretical abundance (12%) of the $^{13}C$ monoisotopic mass, and the $MS^2$ fragments ($m/z$ = 118.07, 132.08, and 145.09).

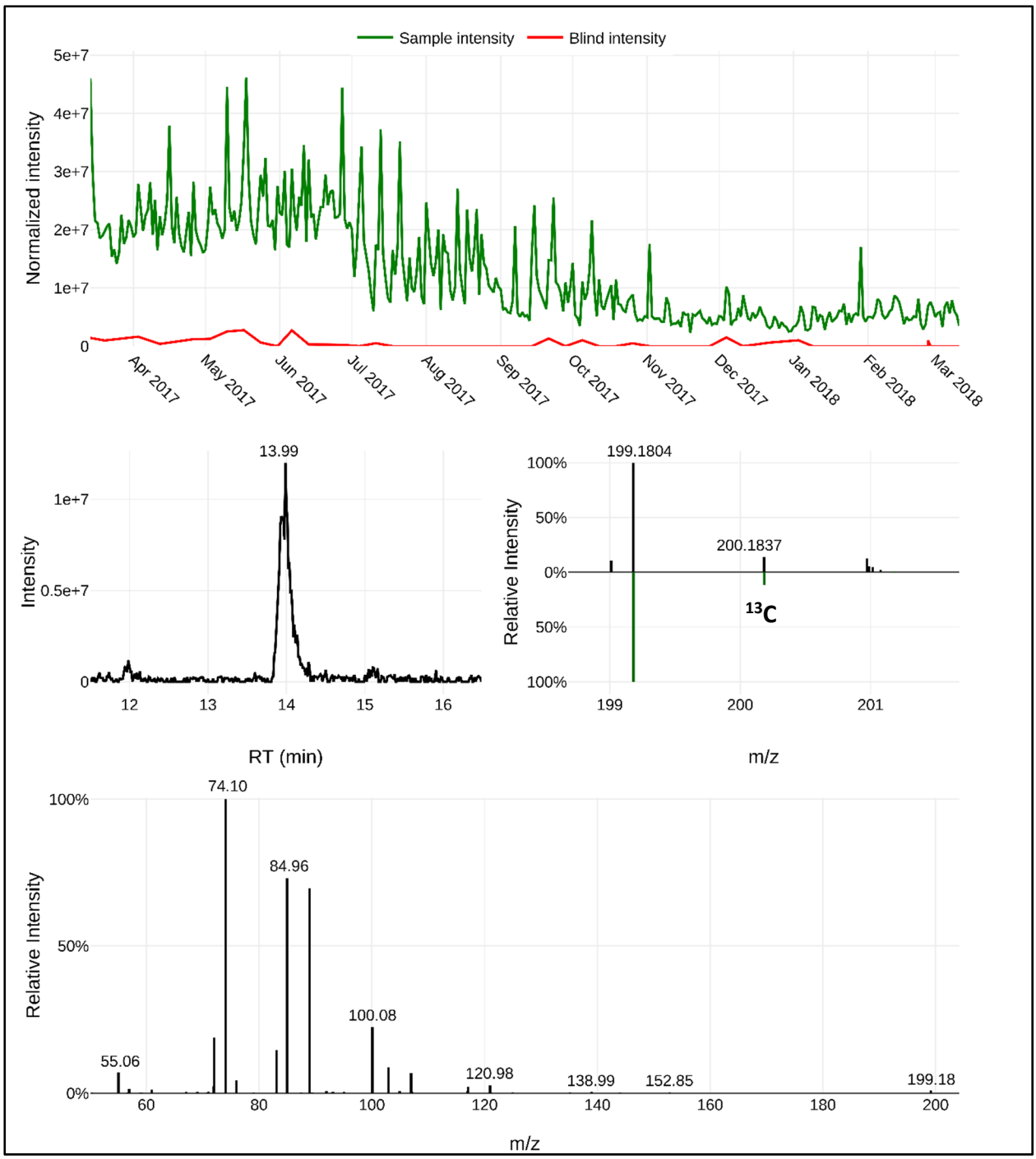



**Figure B88:** Identification of NT199 – level 4.

NT199 was identified to have an unequivocal molecular formula ($C_{11}H_{22}N_2O$) which had a $RT_{avg}$ of 14.0 min. The proposed molecular formula assignment matches the MS spectra ($m/z_{avg}$ = 199.1805 for $[M+H]^+$, $\Delta m$ = 0.18 ppm), the theoretical abundance (11%) of the $^{13}C$ monoisotopic mass, and the $MS^2$ fragments.

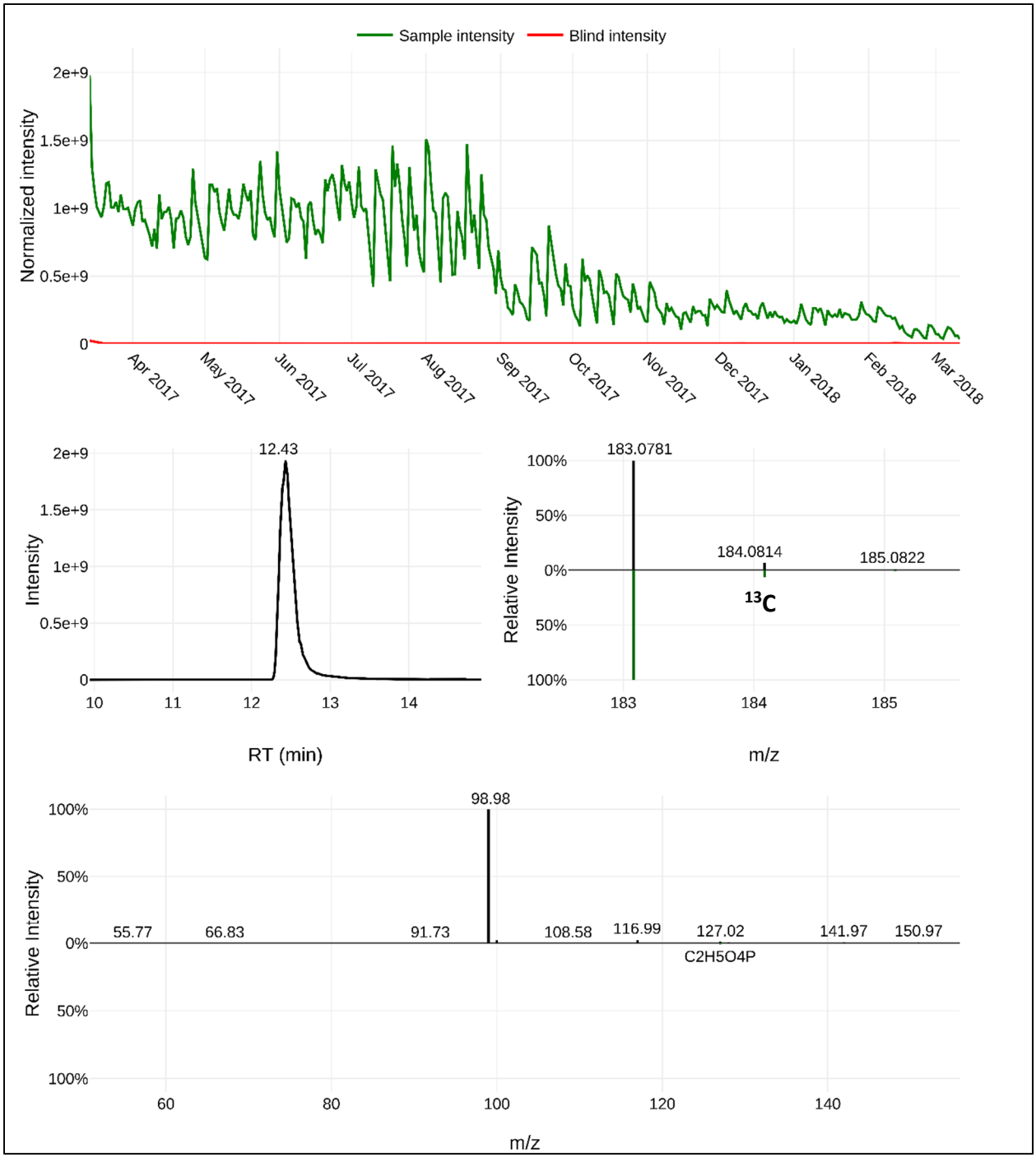



**Figure B89:** Identification of triethyl phosphate – level 1.

Triethyl phosphate ($C_6H_{15}O_4P$) was confirmed using an authentic reference standard which matched the $RT_{avg}$ of 12.4 min, the MS spectra ($m/z_{avg}$ = 183.0781 for $[M+H]^+$, $\Delta m$ = 0.29 ppm), the theoretical abundance (6%) of the $^{13}C$ monoisotopic mass, and the $MS^2$ fragments ($m/z$ = 98.98 and 127.02).

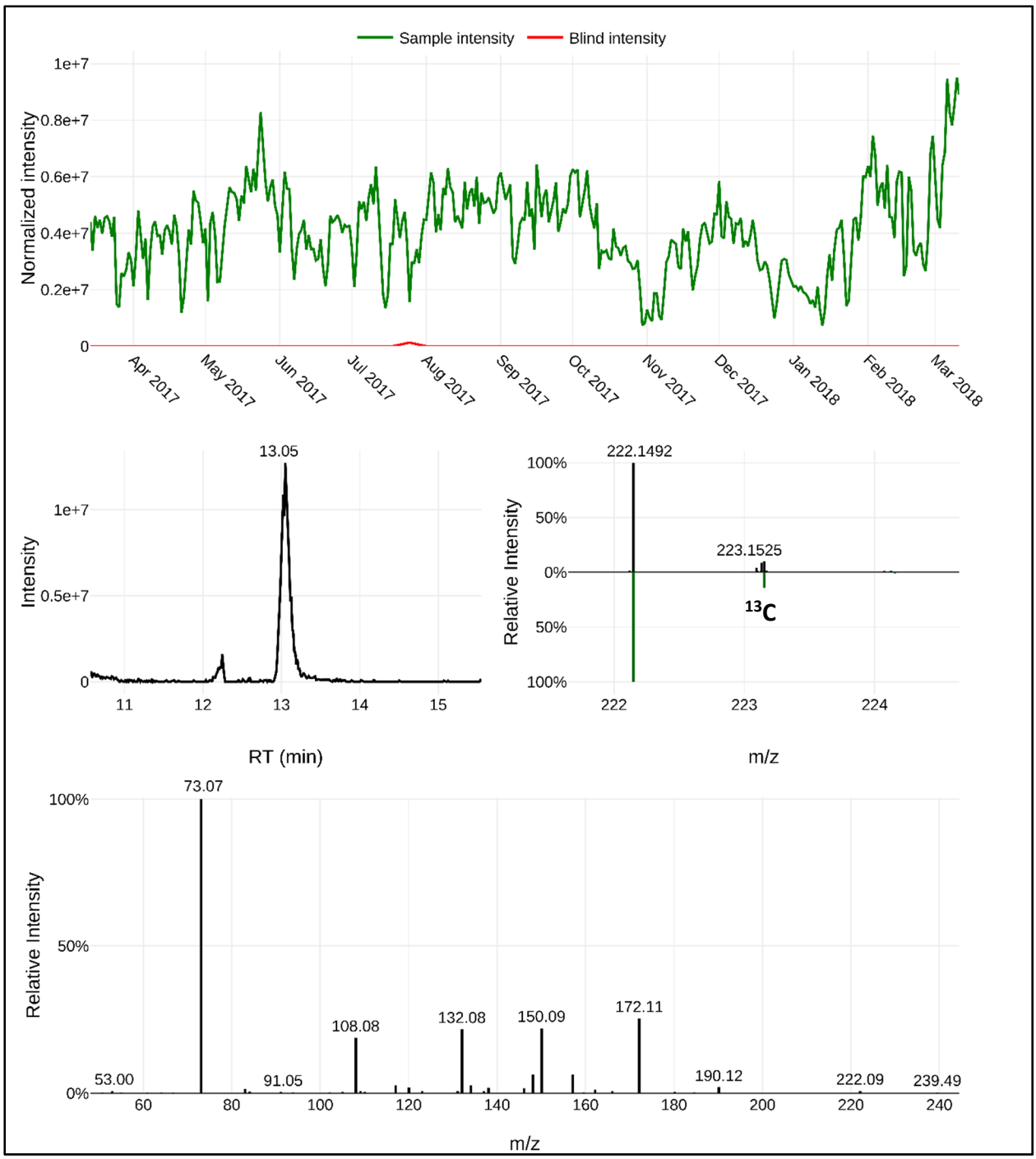



**Figure B90:** Identification of NT222 – level 4.

NT222 was identified to have an unequivocal molecular formula ($C_{13}H_{19}NO_2$) which had a $RT_{avg}$ of 13.1 min. The proposed molecular formula assignment matches the MS spectra ($m/z_{avg}$ = 222.1489 for $[M+H]^+$, $\Delta m$ = 0.17 ppm), the theoretical abundance (13%) of the $^{13}C$ monoisotopic mass, and the $MS^2$ fragments.

# APPENDIX C – Fall Creek Monitoring Station: Using Environmental Covariates to Predict Micropollutant Dynamics and Peak Events in Surface Water Systems

## C.1 Data collection

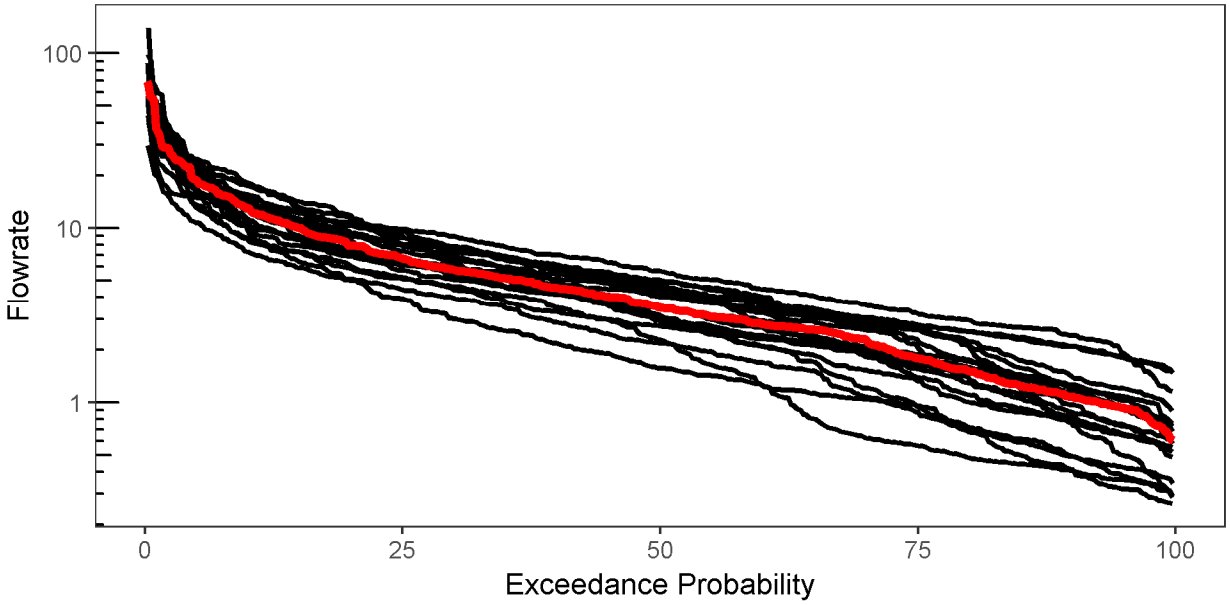



**Figure C1:** Daily stream flowrate ($m^3 \cdot s^{-1}$) duration curves for 2000 – 2018 (black lines) and the study period (red line).

Daily weather data were obtained from the National Oceanic and Atmospheric Administration (NOAA)[11] using the R package *rnoaa*[12] and from the Northeast Regional Climate Center (NRCC).[13] A list of 92 NOAA and 6 NRCC weather stations were identified within a distance of approximately 66 km (4x the watershed radius) from the watershed centroid. A map of the weather stations is provided in **Figure C2**. Spatial interpolations for daily precipitation (*prcp*) data were conducted by first filling missing data from individual weather stations using inverse distance weighting and then applying temporally pooled kriging using the R package *gstat.*[14] The precipitation index (*prcp01*) is a binary factor to reflect days with precipitation; *prcp01* specifically represents days with greater than 1 mm of precipitation as a value of 1, and 0 otherwise. The antecedent dry period (*adp*) for each day was determined by the number of previous days since an average rainfall event greater than 1 mm. All other weather data (minimum temperature, *tmin*; average temperature, *tavg*; maximum temperature, *tmax*; snowfall, *snow*; snow depth, *snwd*; wind speed, *awnd*; leaf wetness, *lwet*; solar radiation, *srad*; and relative humidity, *rhum*) were averaged across the watershed from all available weather stations for each day using Thiessen polygons and the R package *spatstat.*[15] Daily snow melt (*snmt*) was determined by the difference between *snwd* on the preceding day. Streamflow rates (*flow*) were obtained from a U.S. Geological Survey (USGS) stream gage located within 2 km downstream of the Fall Creek Monitoring Station (FCMS).[1] Baseflow rates were calculated using the R package *lfstat*[16] to determine the average baseflow index (*bfi*) for each day. Average daily sewage treatment plant (STP) discharge rates (*stp*) were obtained from one of the two STPs that discharge directly into Fall Creek and normalized by *flow* to determine the contribution of *stp* to *flow*, or the sewage proportion (*pstp*). Data from the other STP could not be obtained. Water quality parameters that are regularly collected at the Cornell Water Filtration Plant through continuous in-line measurements (*pH*;

turbidity, *turb*; and water temperature, *twat*) and daily grab samples ($UV_{254}$ absorbance, *uv*; and alkalinity, *alk*) of the raw water were also included as environmental covariates. Lastly, the age of the water sample (*age*) determined as the number of days between sample collection and retrieval, was also included to account for potential micropollutant degradation during storage.

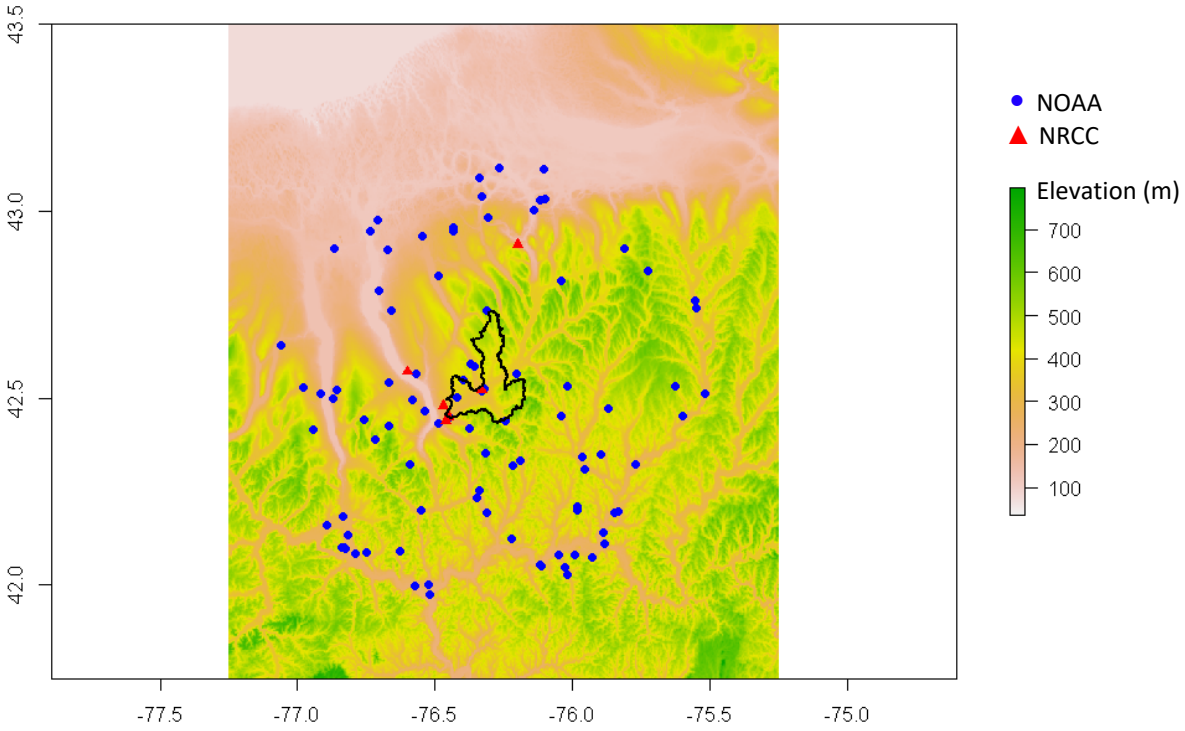



**Figure C2:** Elevations of the Fall Creek watershed (delineated by the black boundary) and surrounding area along with locations of the 92 NOAA weather stations (blue circles) and 6 NRCC weather stations (red triangles).

## C.2 Data processing

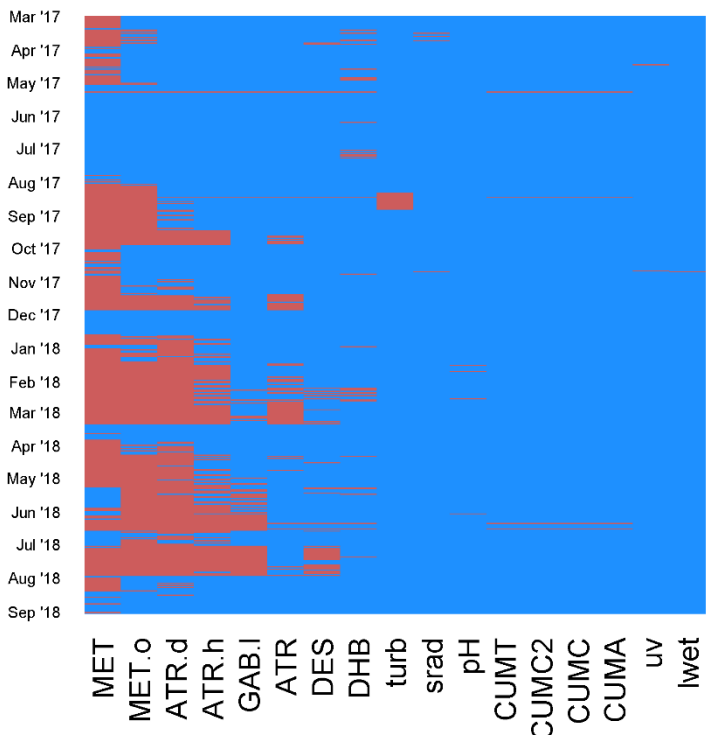



**Figure C3:** Missing values (red) throughout the study period.
(MET = 63.2% missing; MET.o = 47.5%; ATR.d = 43.3%; ATR.h = 28.4%; GAB.l = 12.1%; ATR = 11.8%; DES = 8.0%; DHB = 6.2%; turb = 2.9%; srad = 0.7%; pH = 0.7%; CUMC = 0.7%; CUMA = 0.7%; CUMT = 0.7%; uv = 0.4%; and lwet = 0.2%).

Complete temporal profiles for the representative individual micropollutant concentrations (**Figures C4 – C11**), cumulative metrics of overall micropollutant contamination (**Figures C12 – C15**), and environmental covariates (**Figures C16 – C38** are provided below. The black line represents the measured data. Red segments represent the standard deviation of the five imputed results. For the temporal profiles of micropollutant concentrations and cumulative metrics of overall micropollutant contamination, the blue line represents the peak event limit (highest quartile of values) and the green shaded areas represent the apparent growing seasons.

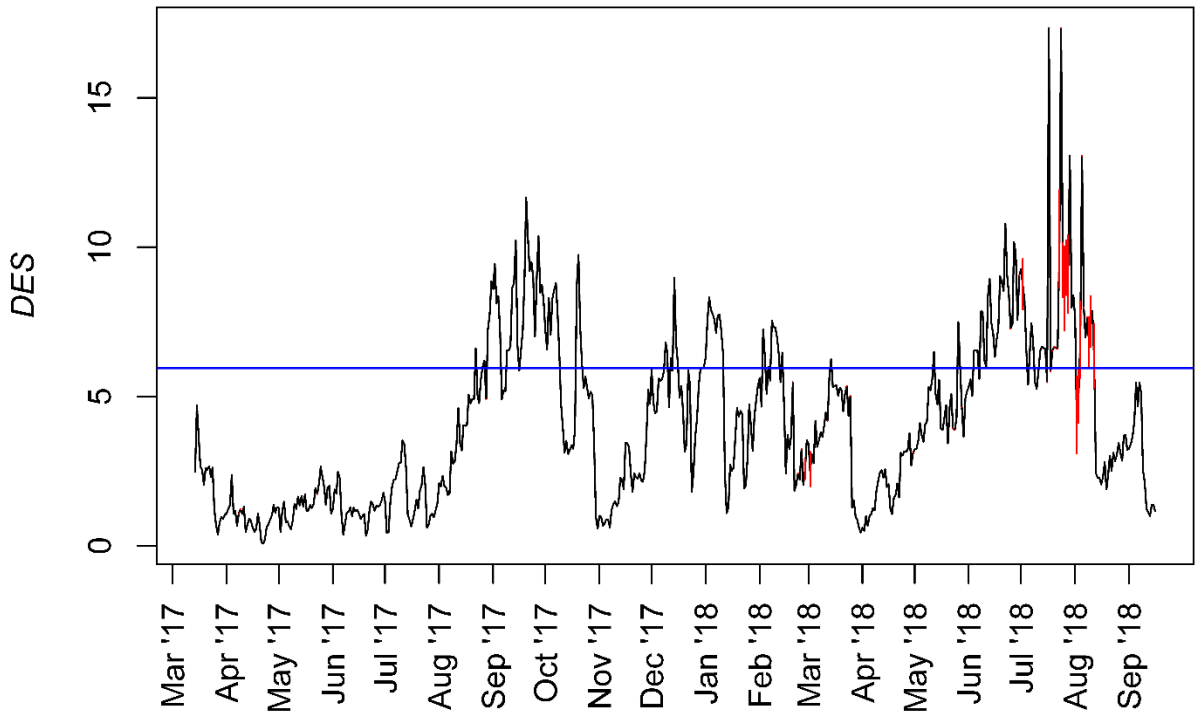



**Figure C4:** Temporal profile of desvenlafaxine (DES) concentration in ng·L$^{-1}$.

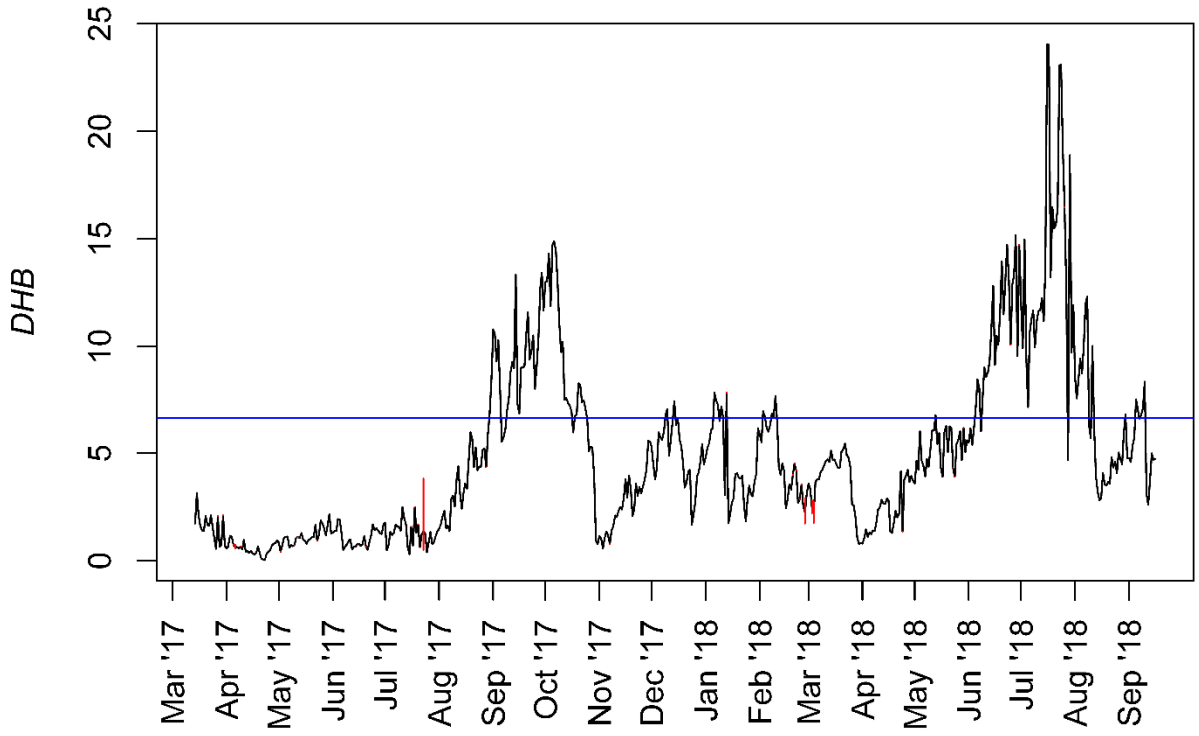



**Figure C5:** Temporal profile of rac-threo-dihydrobupropion (DHB) concentration in ng·L$^{-1}$.

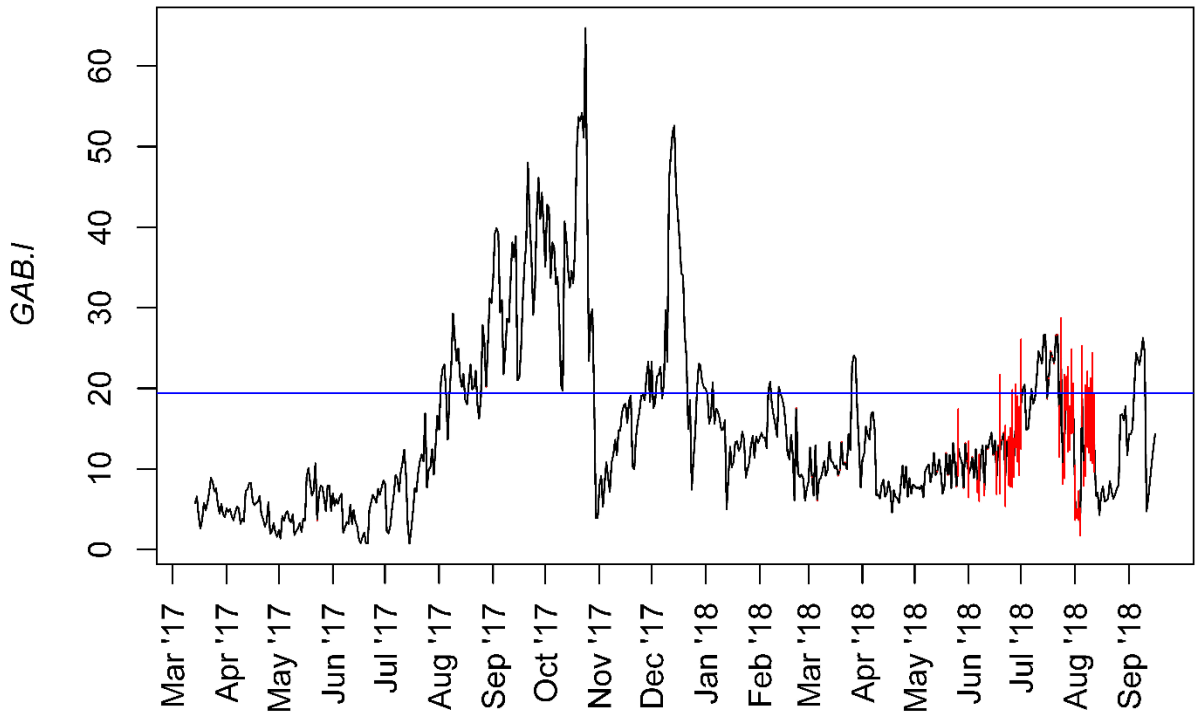



**Figure C6:** Temporal profile of gabapentin-lactam (GAB.l) concentration in $ng \cdot L^{-1}$.

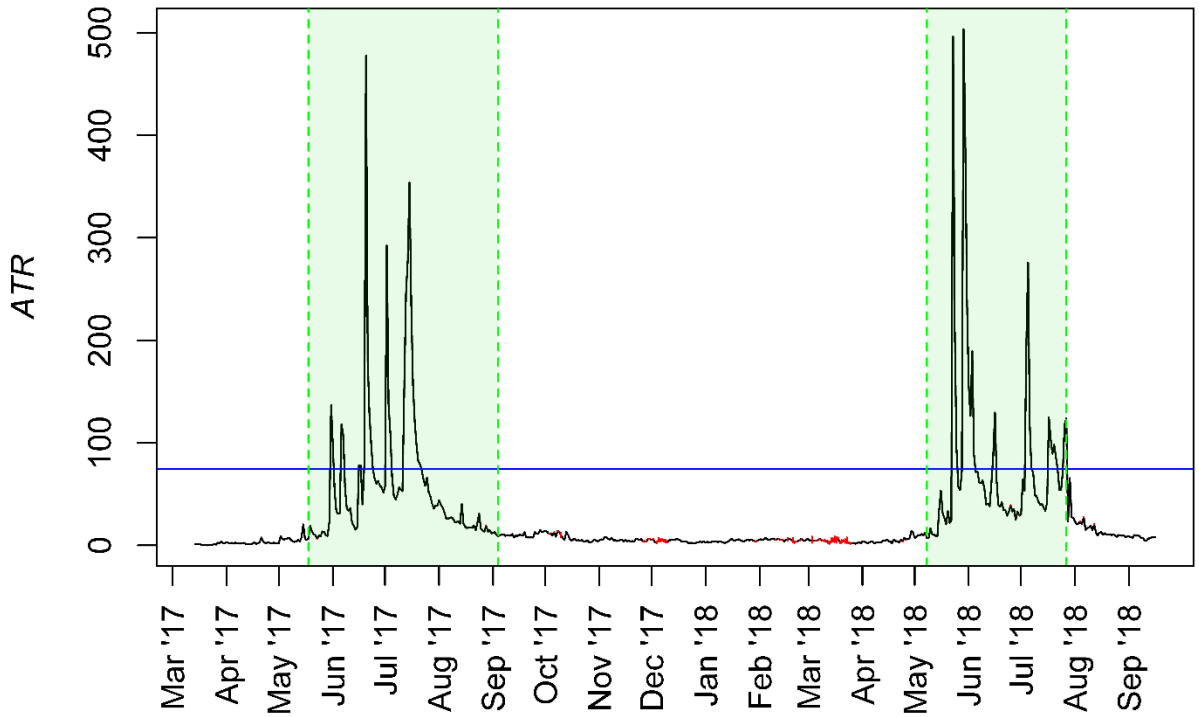



**Figure C7:** Temporal profile of atrazine (ATR) concentration in $ng \cdot L^{-1}$.

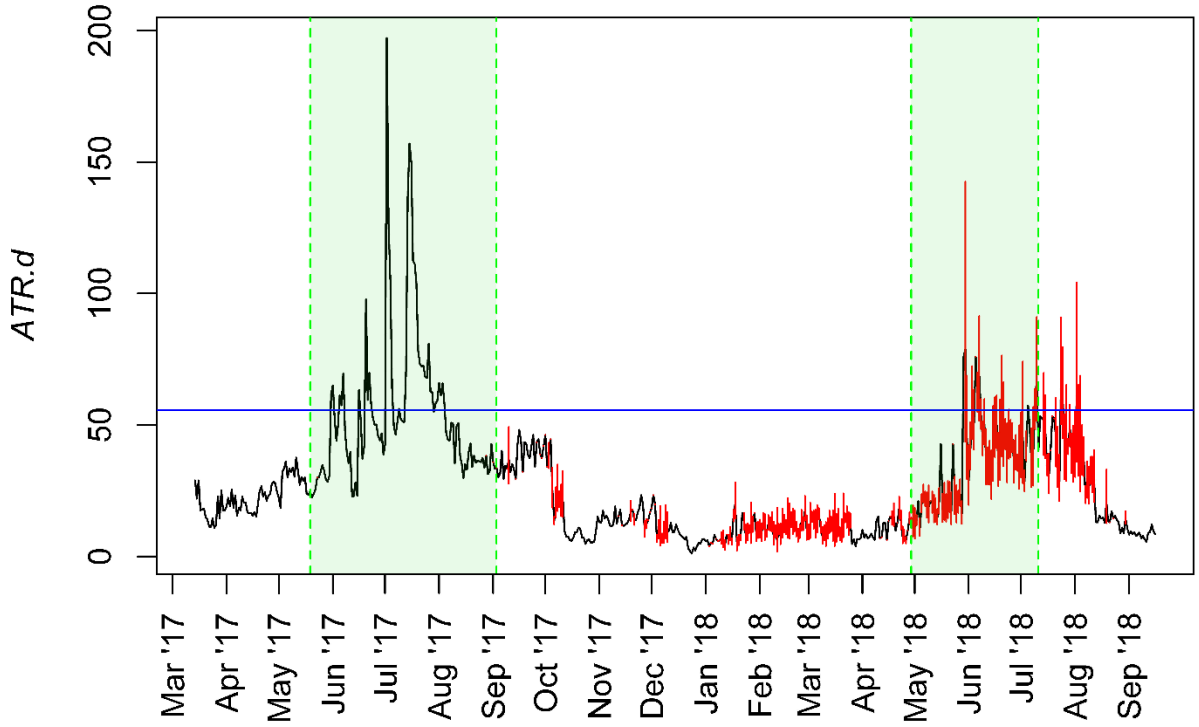



**Figure C8:** Temporal profile of atrazine-desethyl (ATR.d) concentration in ng·L$^{-1}$.

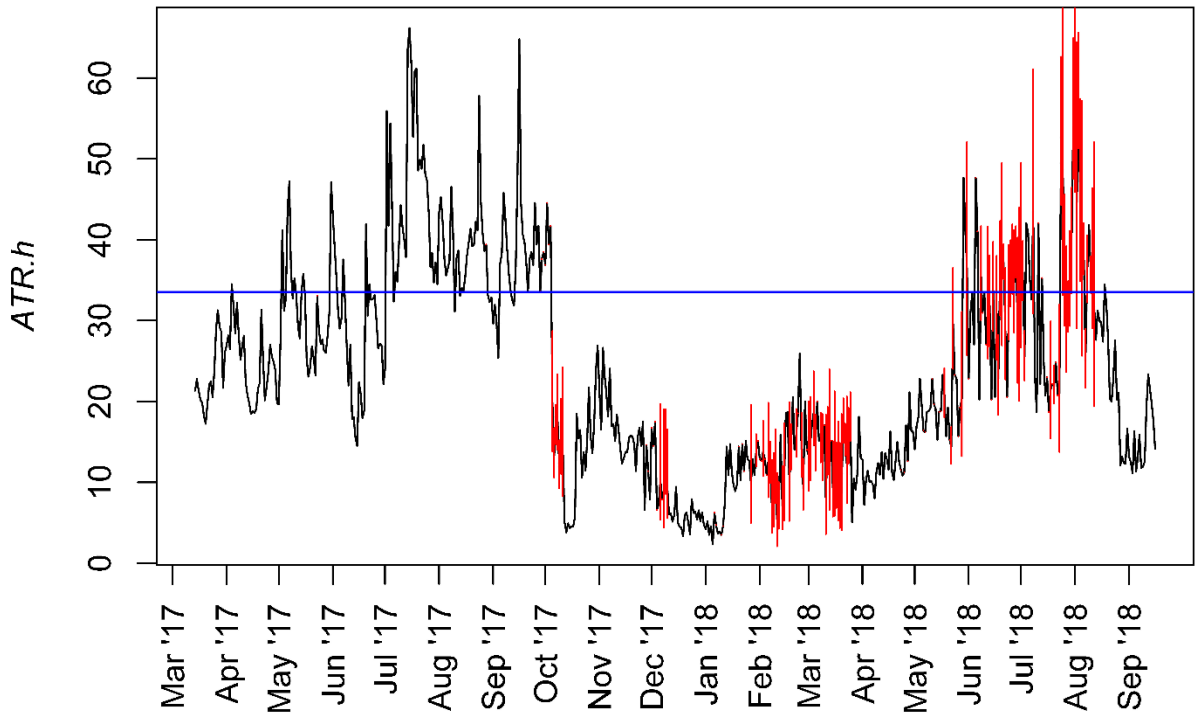



**Figure C9:** Temporal profile of atrazine-2-hydroxy (ATR.h) concentration in ng·L$^{-1}$.

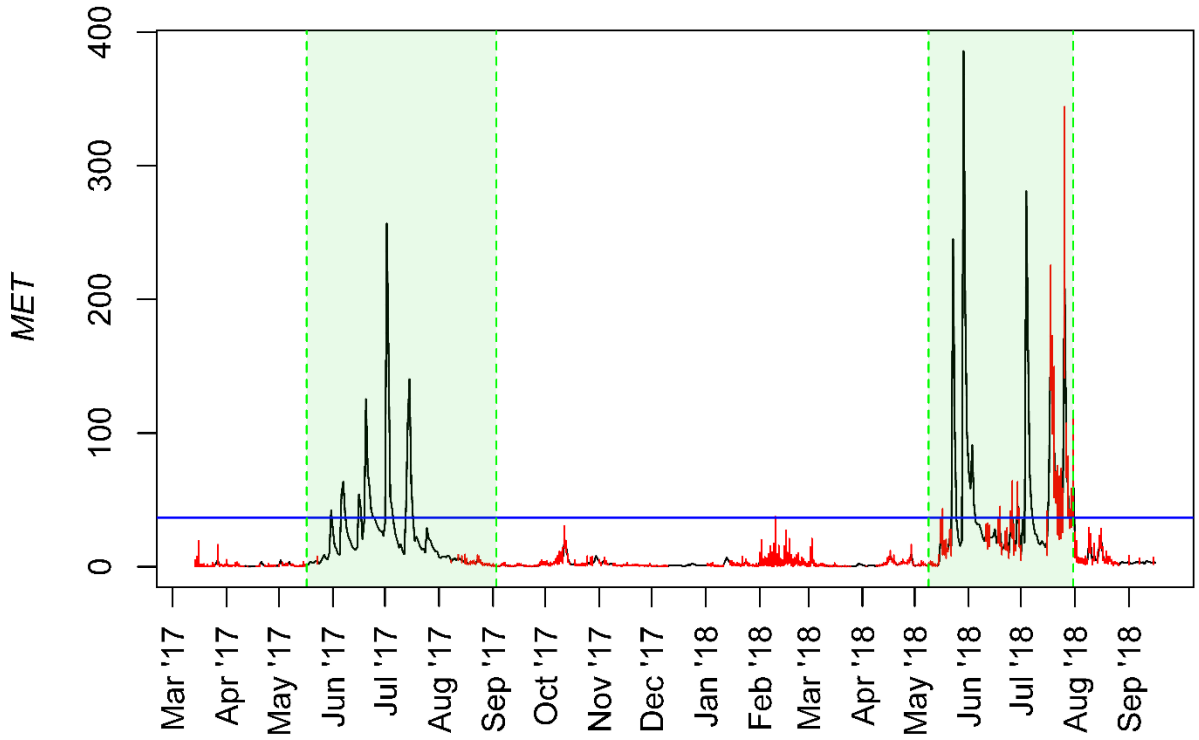



**Figure C10:** Temporal profile of metolachlor (MET) concentration in ng·L$^{-1}$.

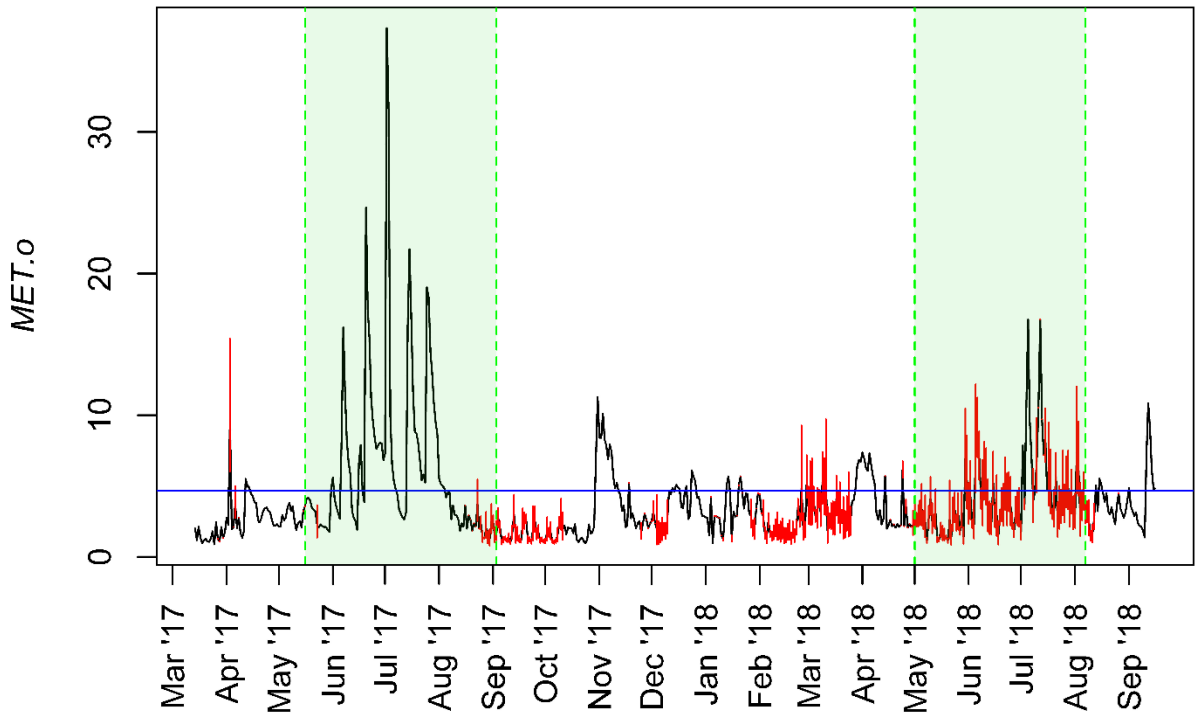



**Figure C11:** Temporal profile of metolachlor-OXA (MET.o) concentration in ng·L$^{-1}$.

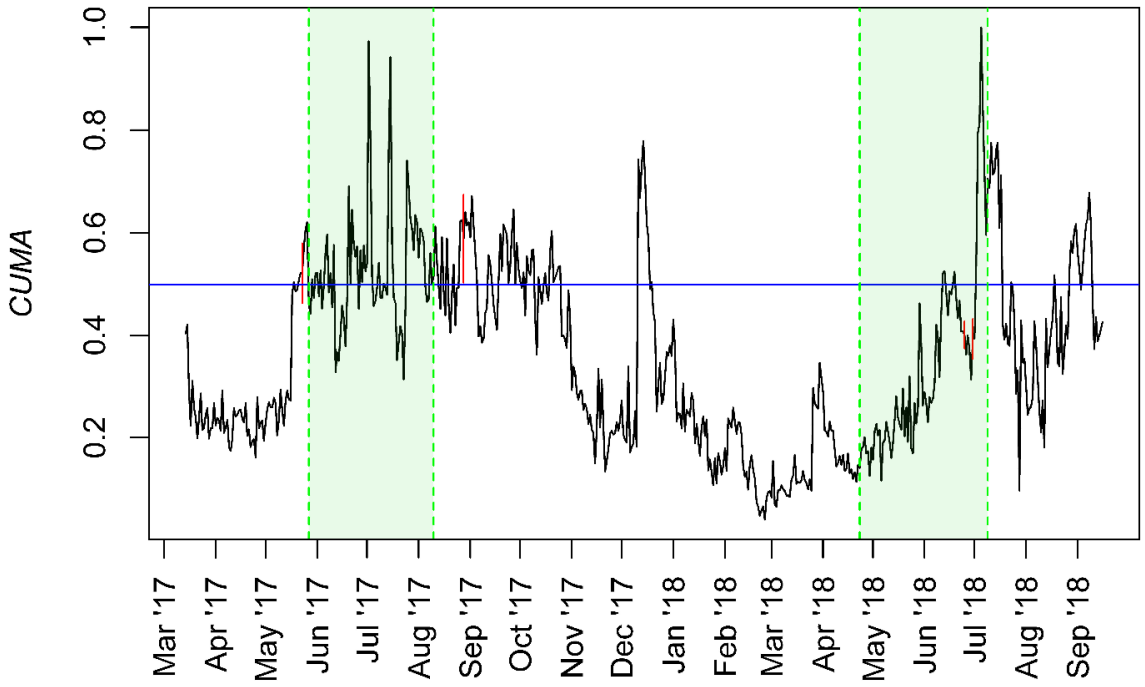



**Figure C12:** Temporal profile of cumulative normalized intensity (CUMA), which represents the sum of 90 previously reported micropollutants (*i.e.*, temporal abundance profiles of all 90 detected micropollutants were normalized to the maximum abundance in each profile and then summed together).

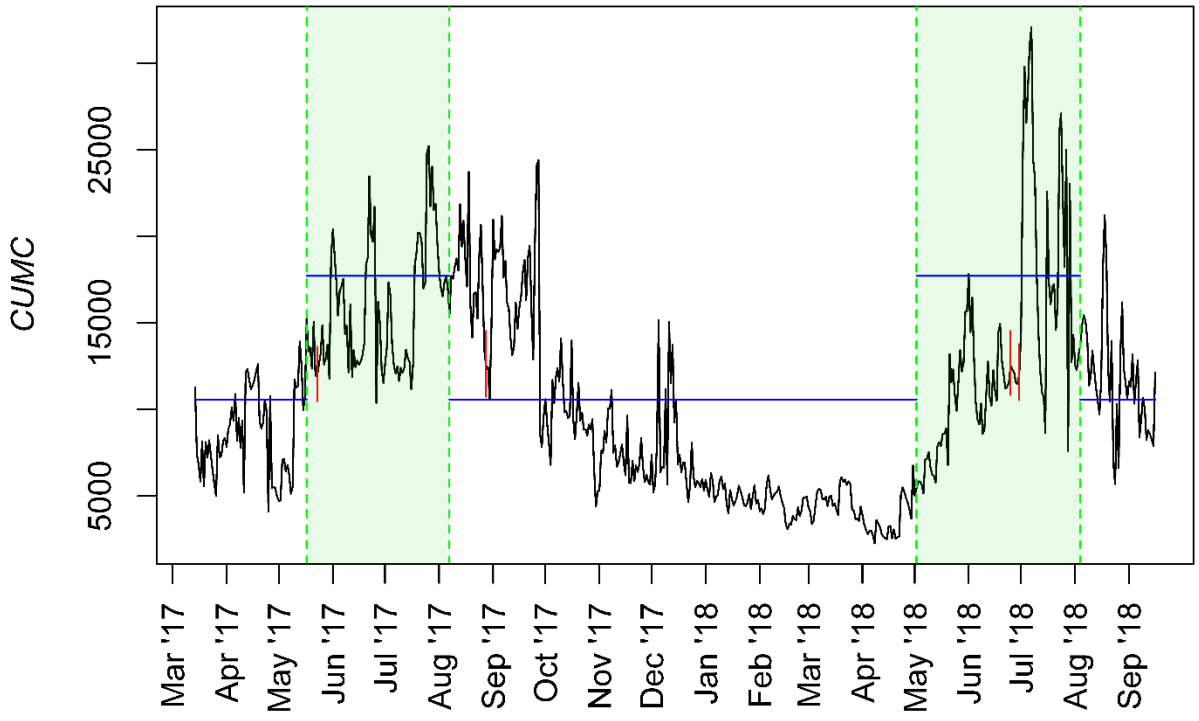



**Figure C13:** Temporal profile of cumulative estimated concentration (CUMC) in $ng \cdot L^{-1}$, which represents 42 micropollutants that we previously identified with level 1 confidence.

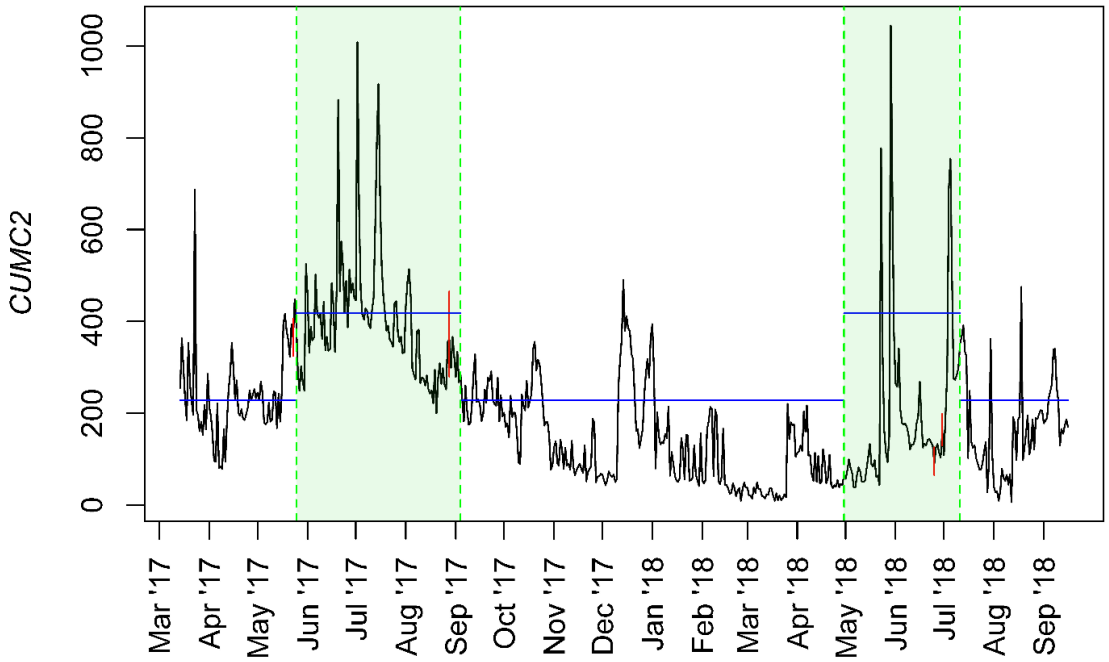



**Figure C14:** Temporal profile of cumulative estimated concentration (CUMC2) in $ng \cdot L^{-1}$, which represents the 32 micropollutants that were consistently measured at levels below 1 $\mu g \cdot L^{-1}$ over the duration of the study period.

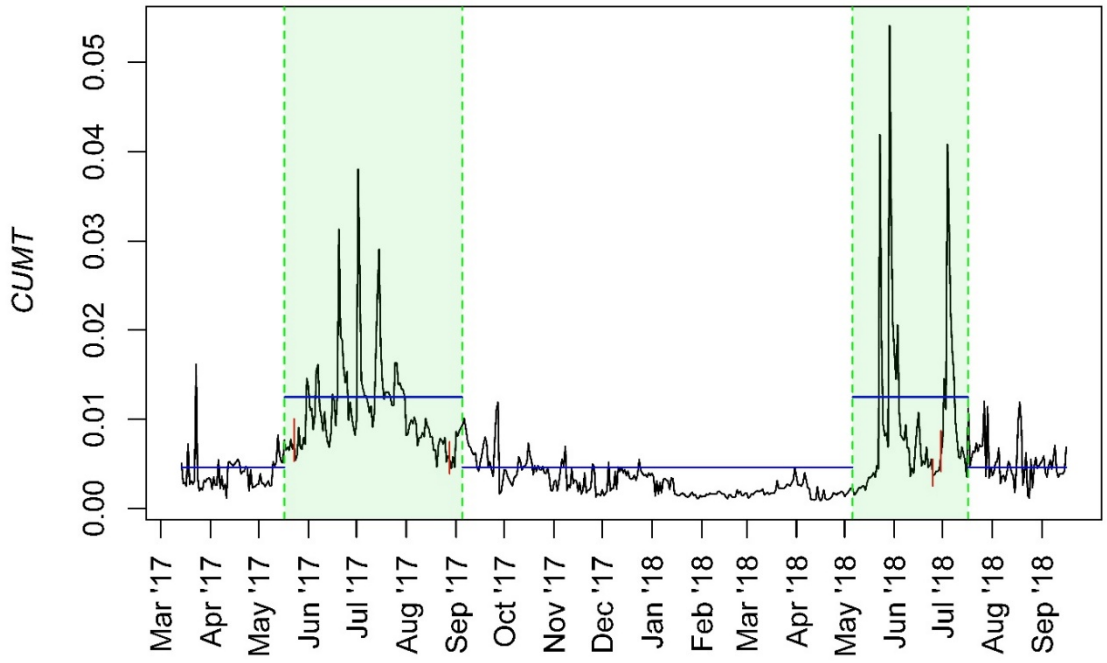



**Figure C15:** Temporal profile of cumulative estimated toxicity (CUMT) as exposure-activity ratio (EAR), which represents the 17 micropollutants that were identified with level 1 confidence and for which toxicity data was available.

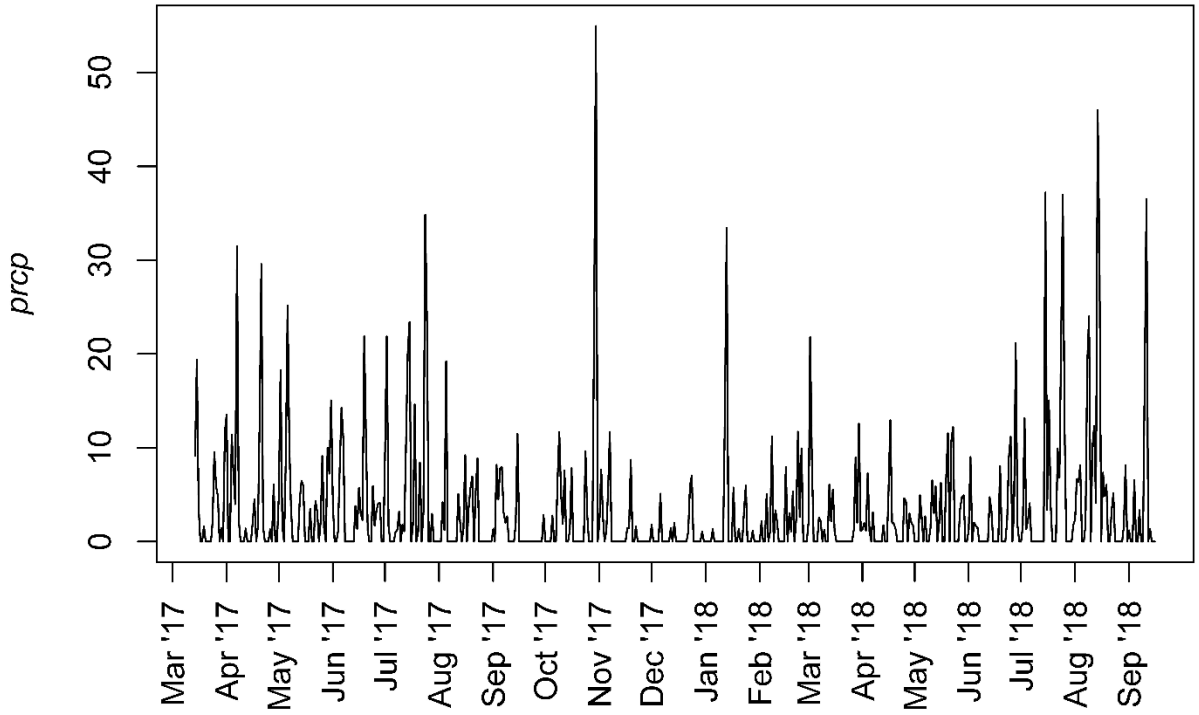



**Figure C16:** Temporal profile of total daily precipitation (*prcp*) in mm.

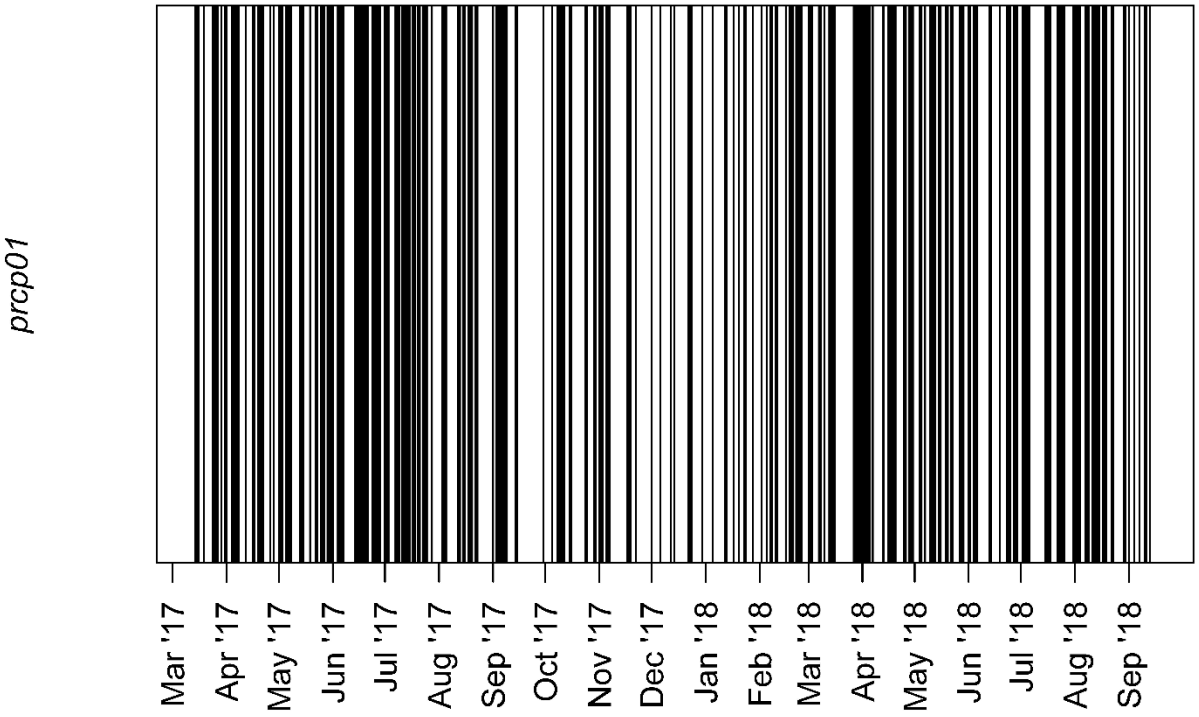



**Figure C17:** Temporal profile of daily precipitation index (*prcp01* – binary value describing days with precipitation greater than 1 mm).

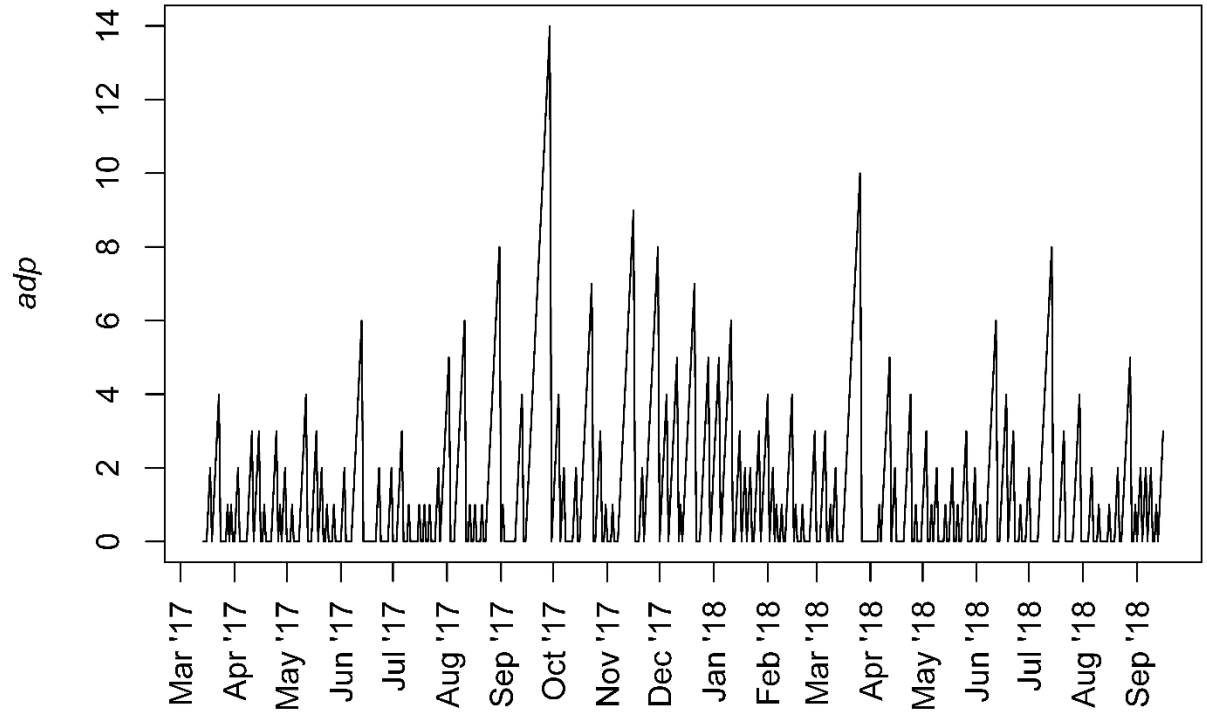



**Figure C18:** Temporal profile of daily antecedent dry period (*adp*) in days.

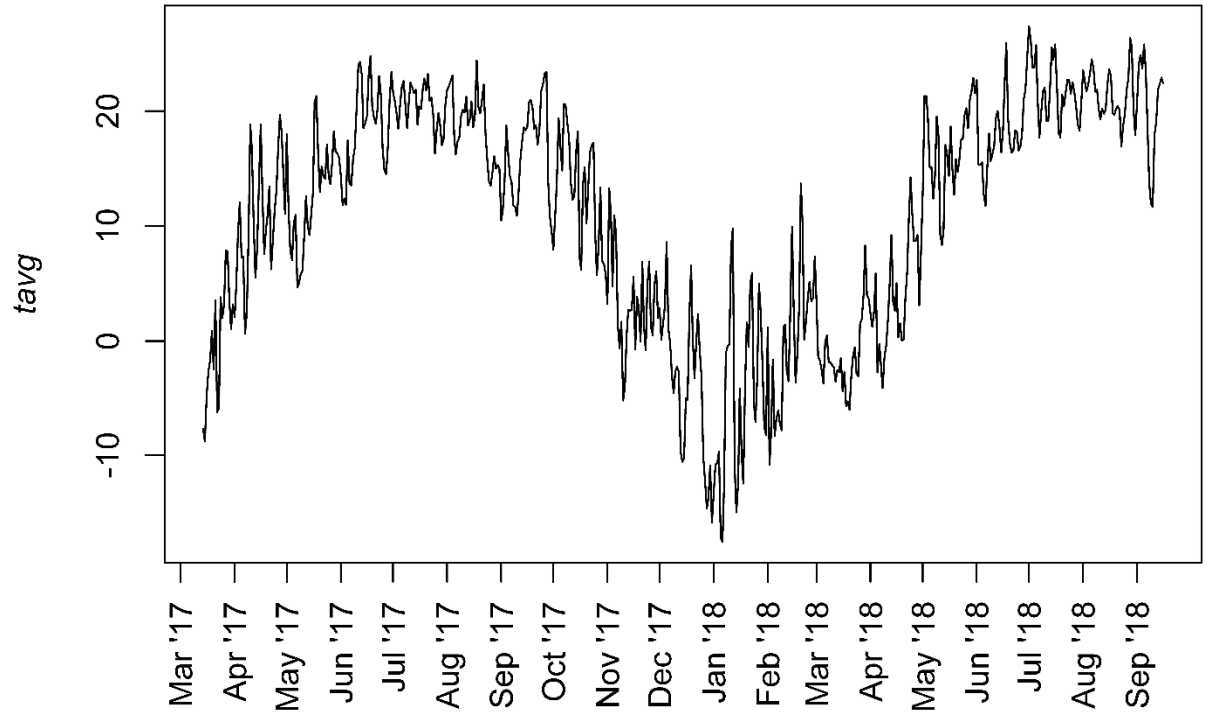



**Figure C19:** Temporal profile of average daily air temperature (*tavg*) in °C.

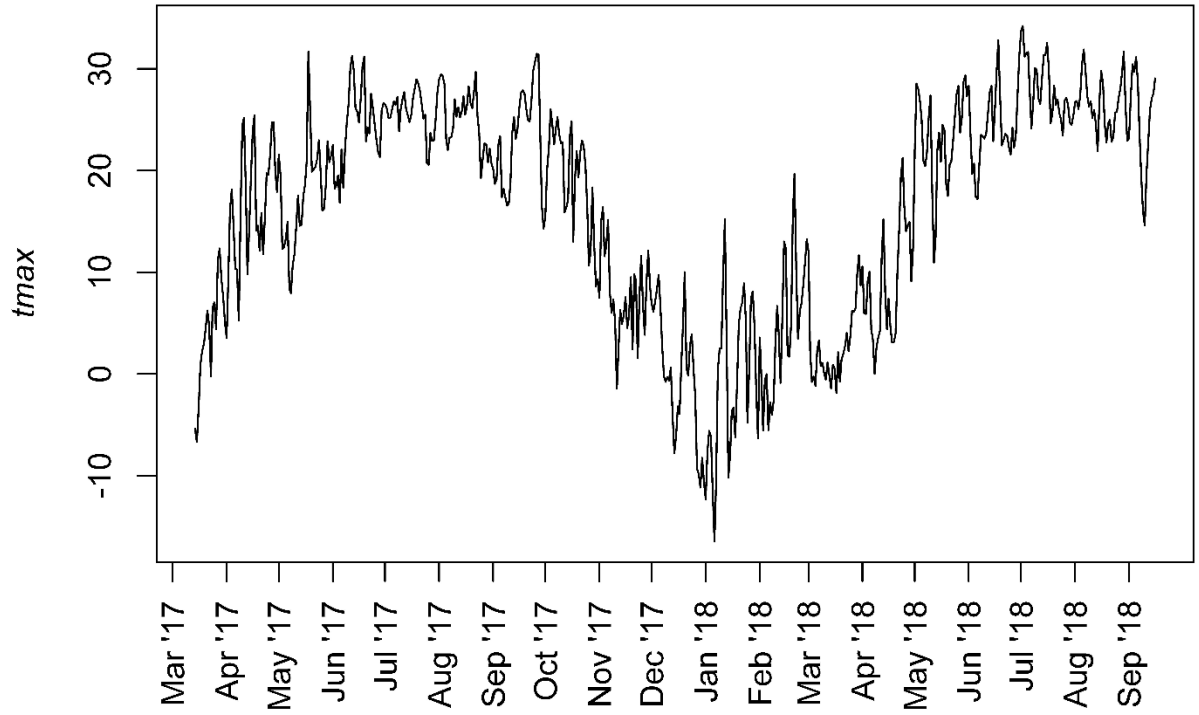



**Figure C20:** Temporal profile of average maximum daily air temperature (*tmax*) in °C.

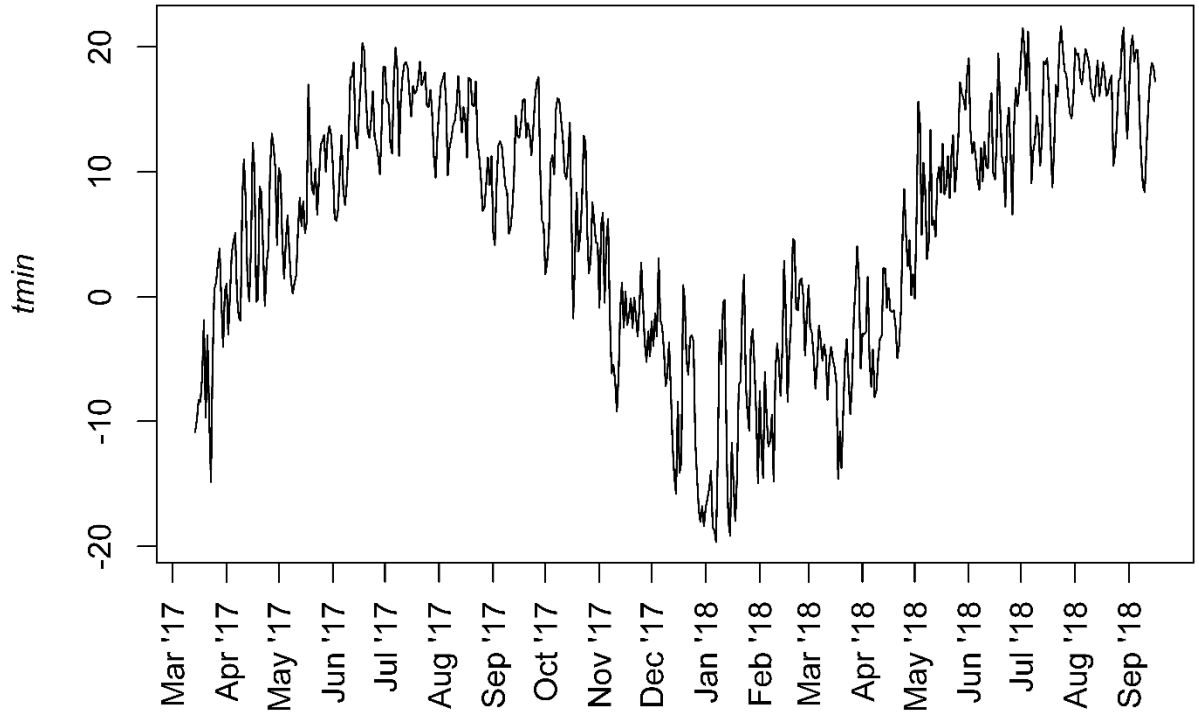



**Figure C21:** Temporal profile of average minimum daily air temperature (*tmin*) in °C.

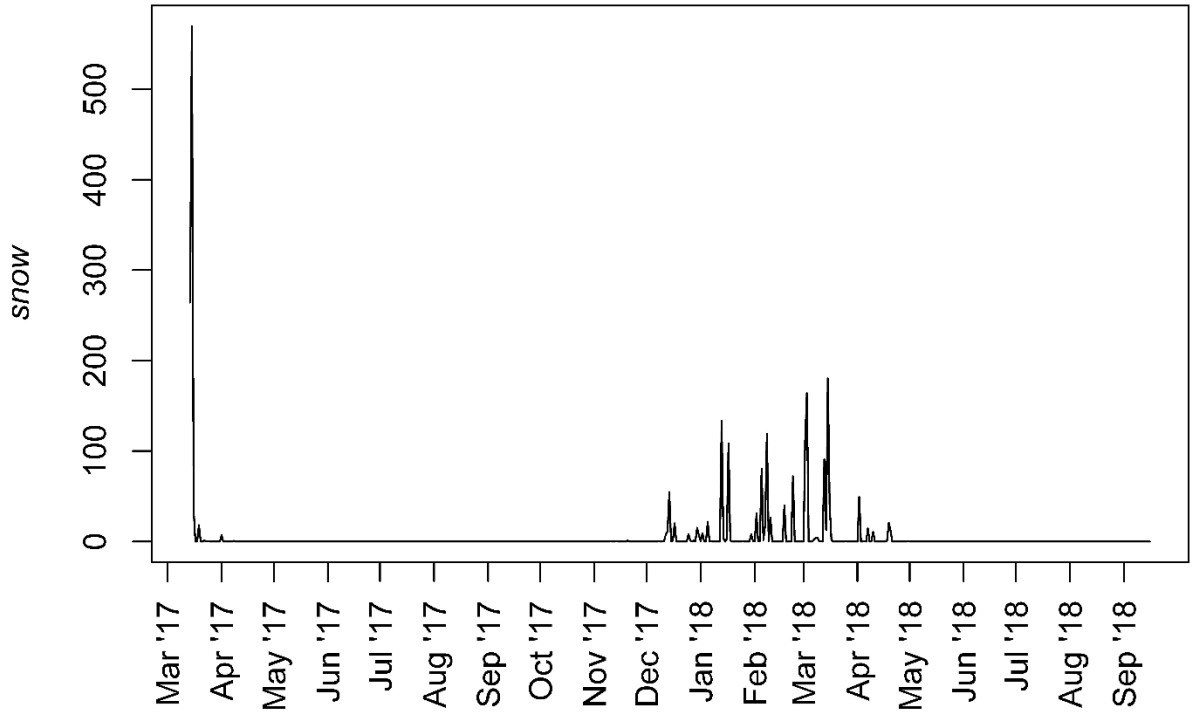



**Figure C22:** Temporal profile of average daily snowfall (*snow*) in mm.

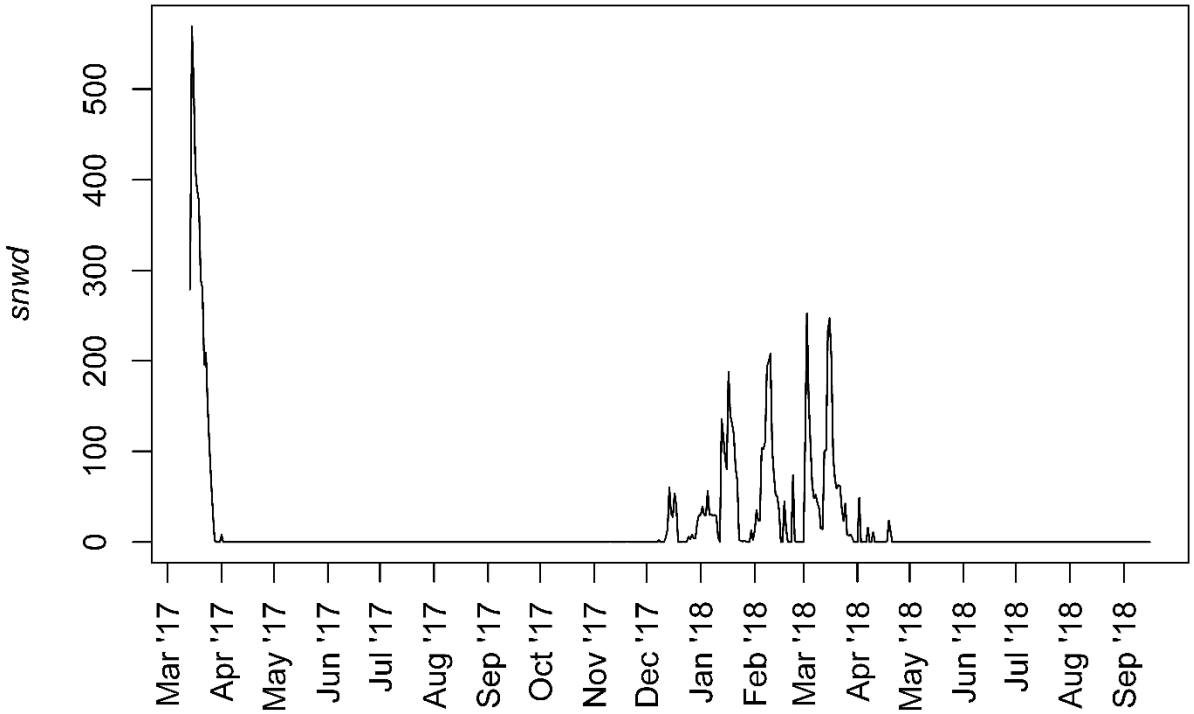



**Figure C23:** Temporal profile of average daily snow depth (*snwd*) in mm.

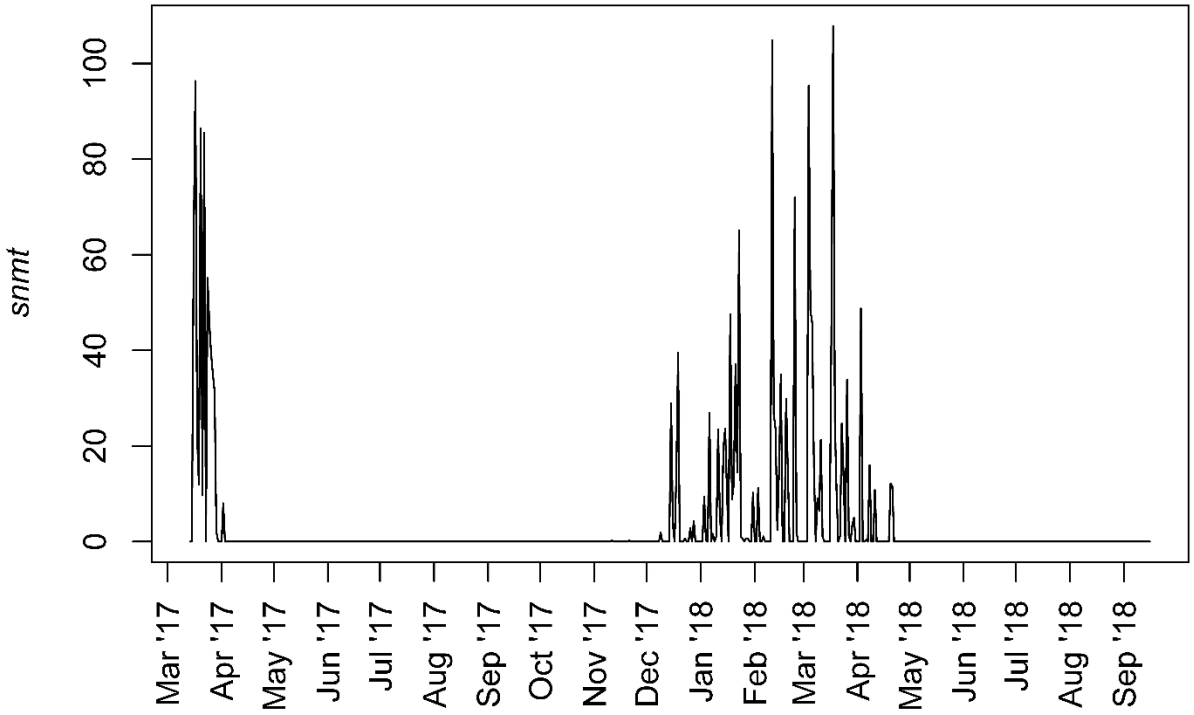



**Figure C24:** Temporal profile of average daily snow melt (*snmt*) in mm.

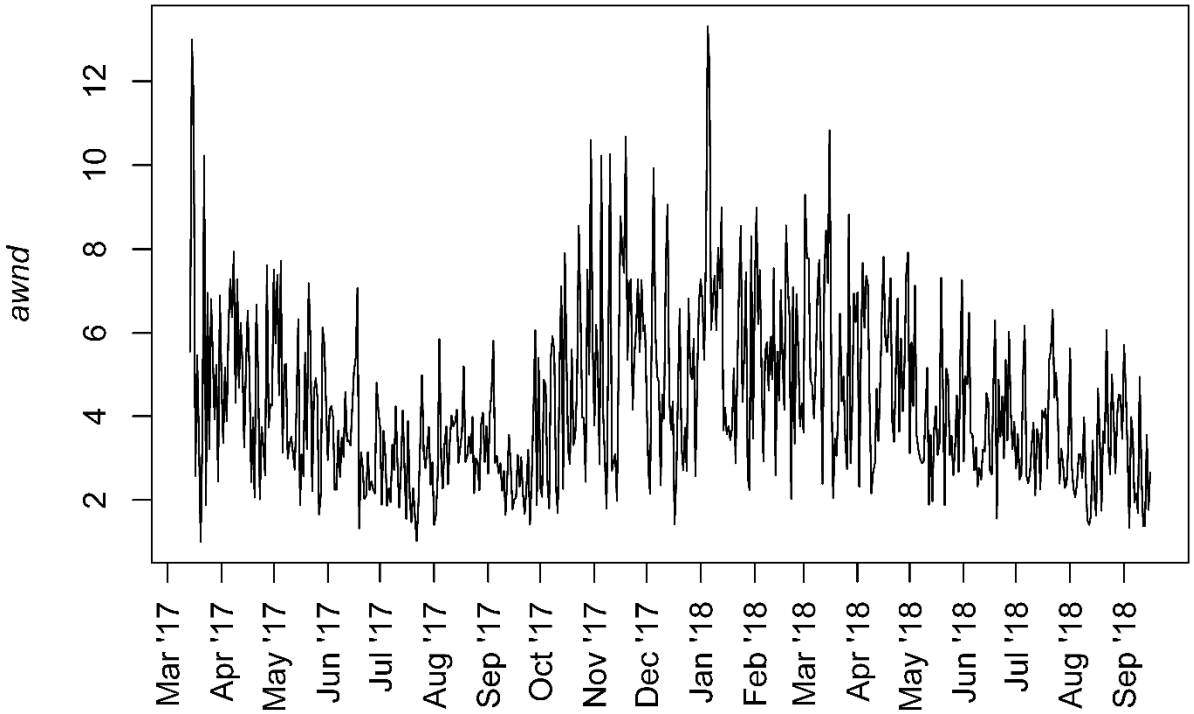



**Figure C25:** Temporal profile of average daily wind speed (*awnd*) in $m \cdot s^{-1}$.

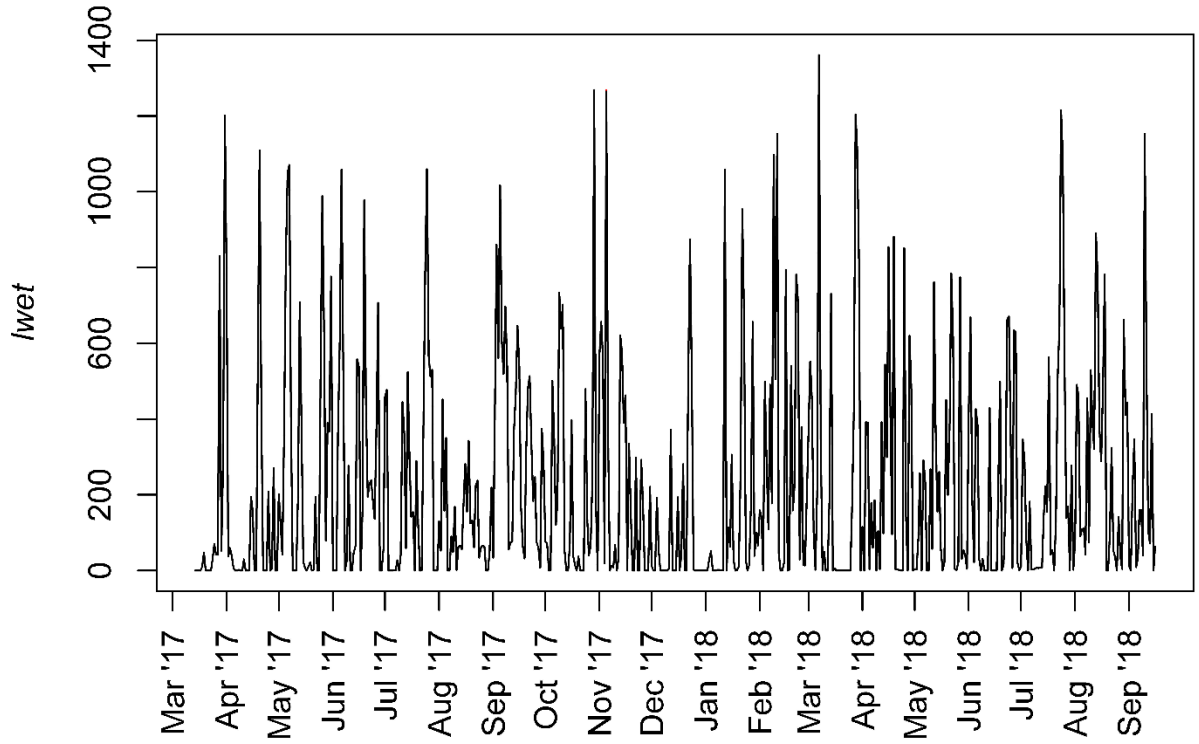



**Figure C26:** Temporal profile of total daily leaf wetness (*lwet)* in min.

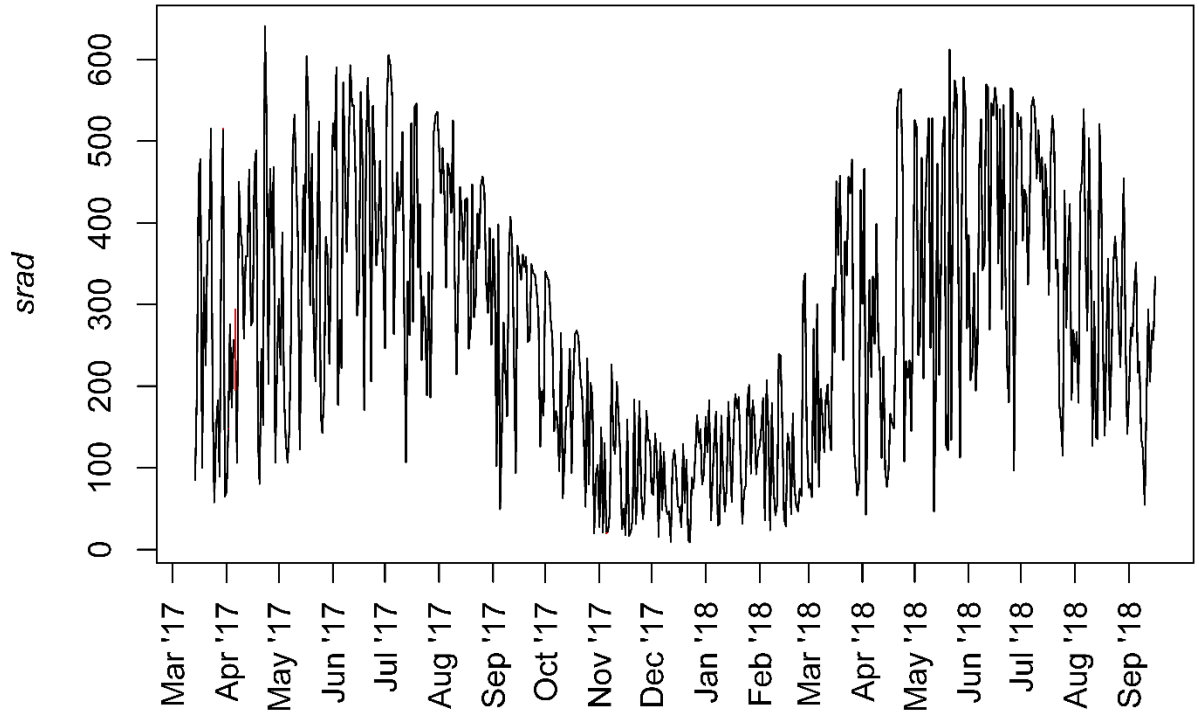



**Figure C27:** Temporal profile of total daily solar radiation (*srad*) in langleys.

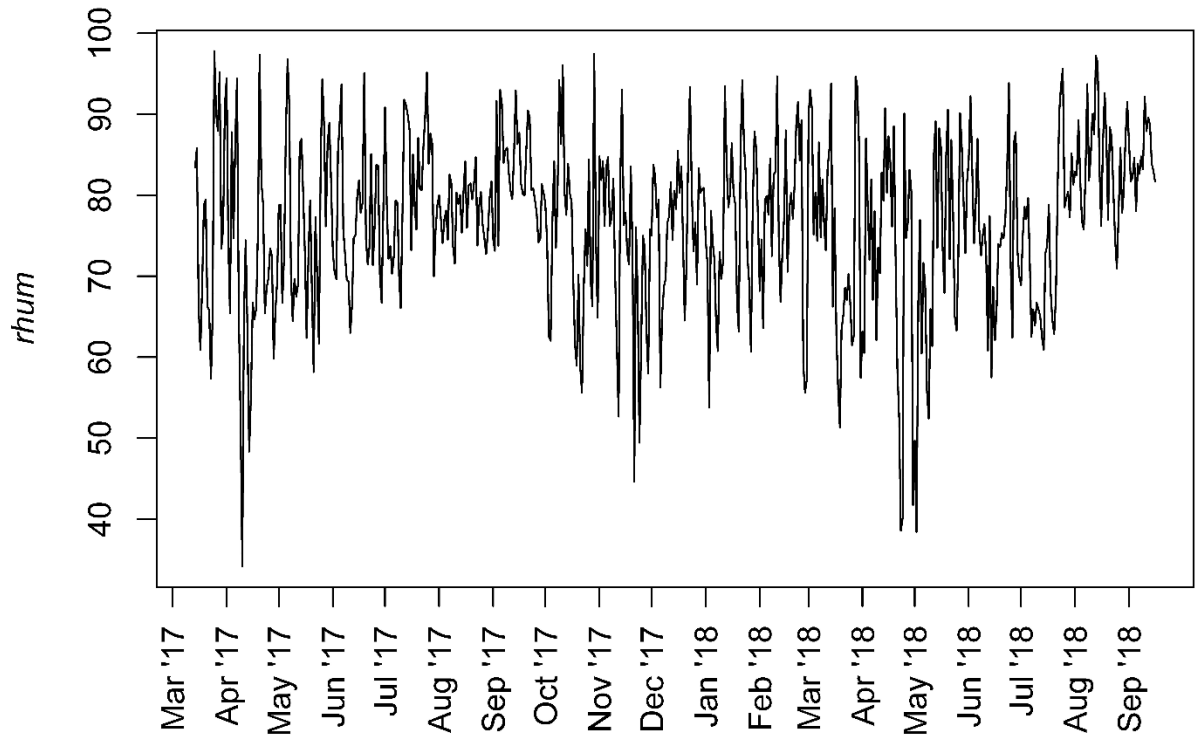



**Figure C28:** Temporal profile of average daily relative humidity (*rhum*) in %.

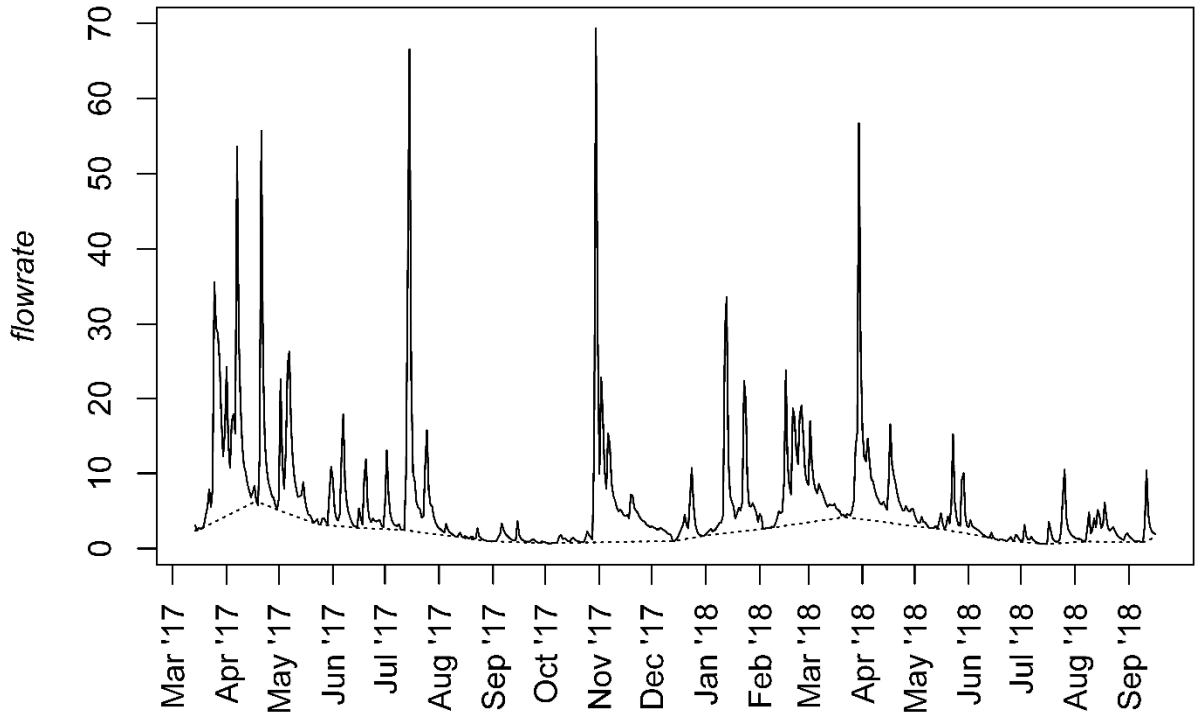



**Figure C29:** Temporal profile of average daily streamflow rate (*flow*; solid line*)* and baseflow rate (dotted line) in $m^3 \cdot s^{-1}$.

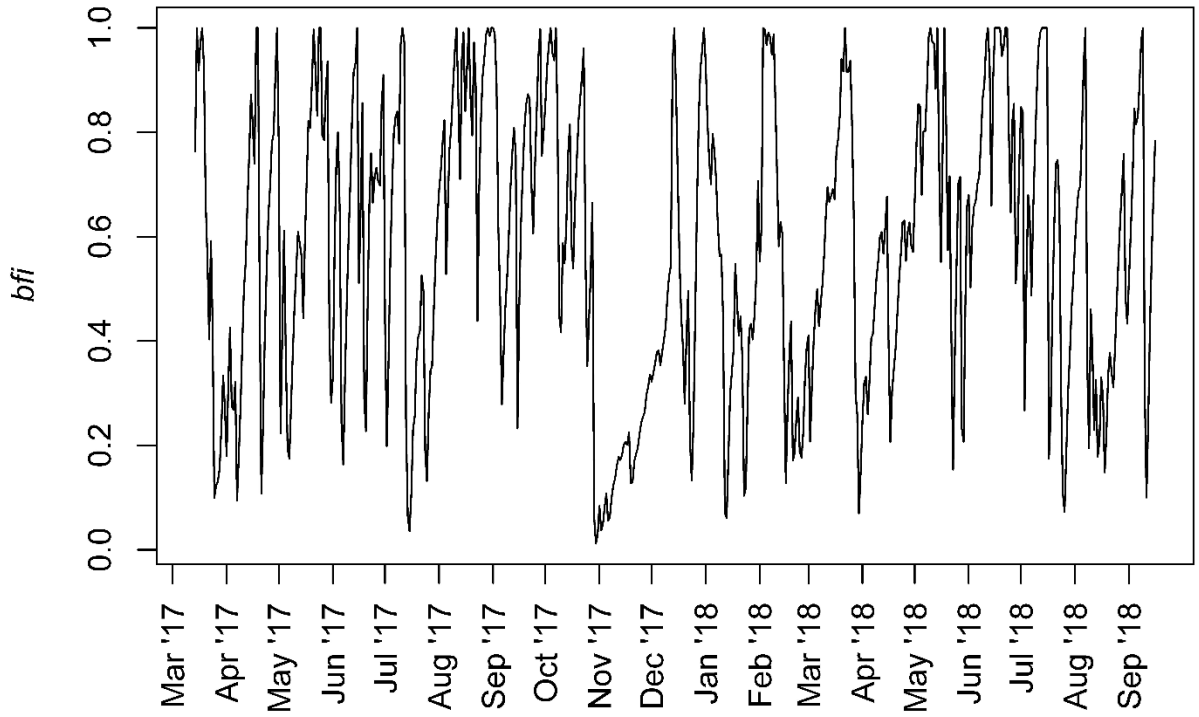



**Figure C30:** Temporal profile of average daily baseflow index (*bfi*).

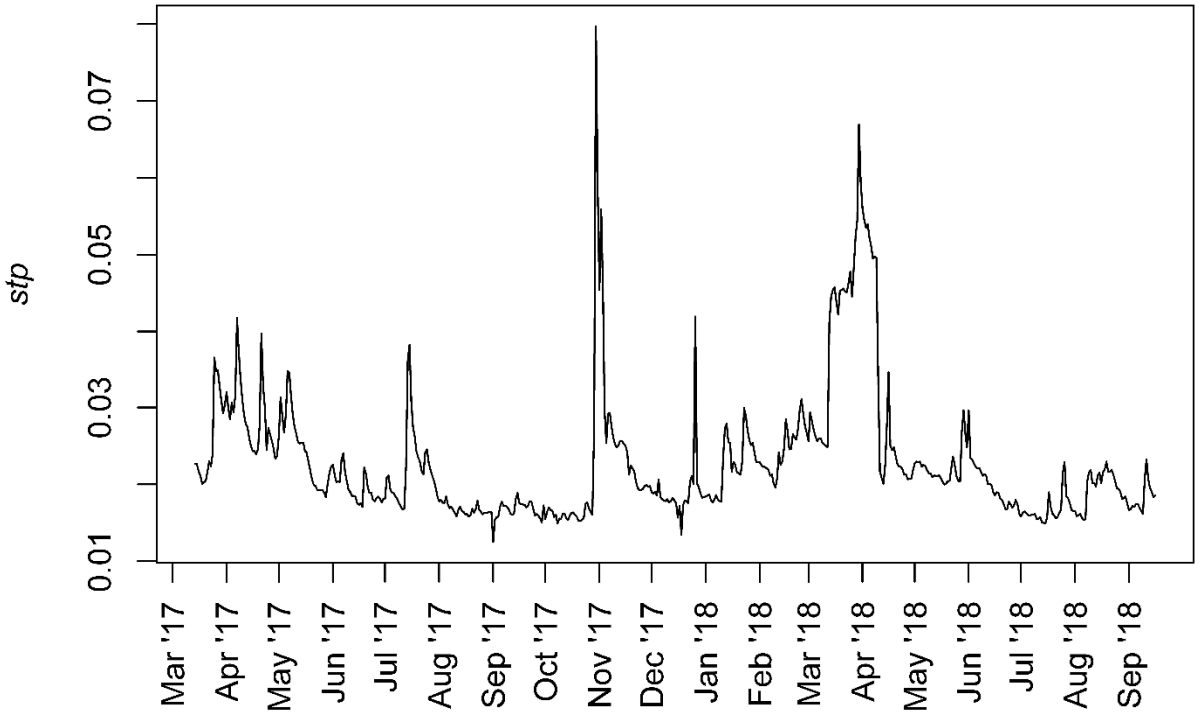



**Figure C31:** Temporal profile of average daily sewage treatment plant discharge (*stp*) in $m^3 \cdot s^{-1}$.

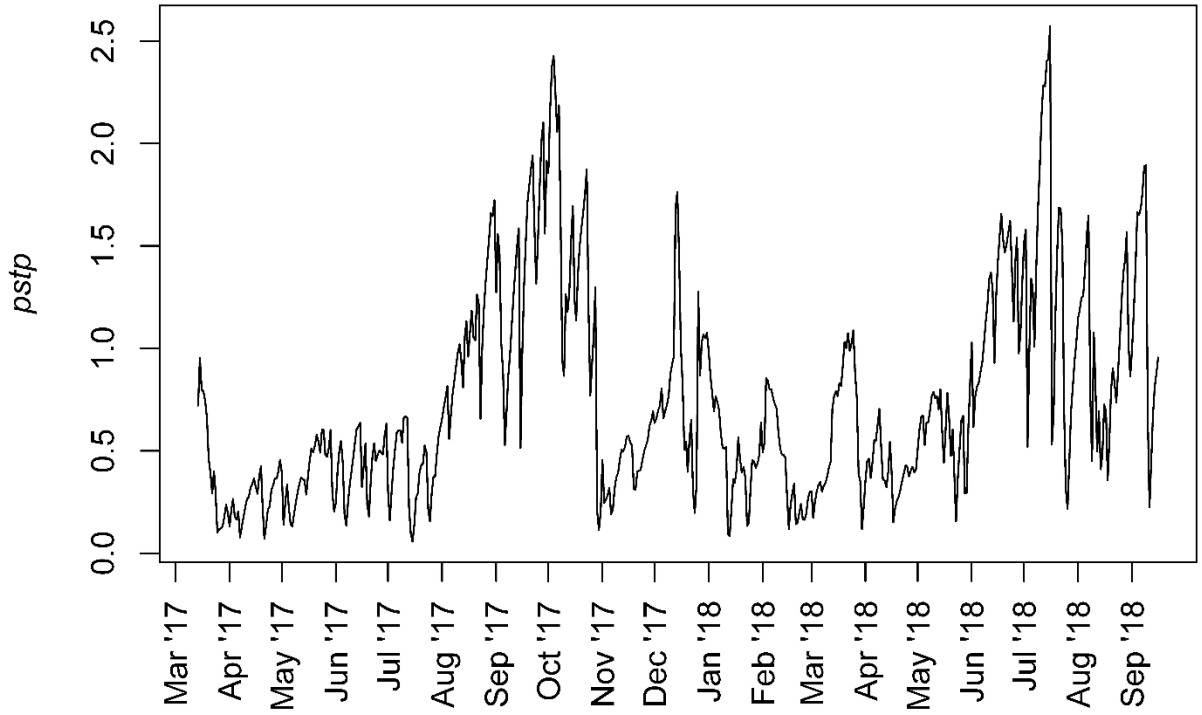



**Figure C32:** Temporal profile of average daily sewage proportion (*pstp*) in %.

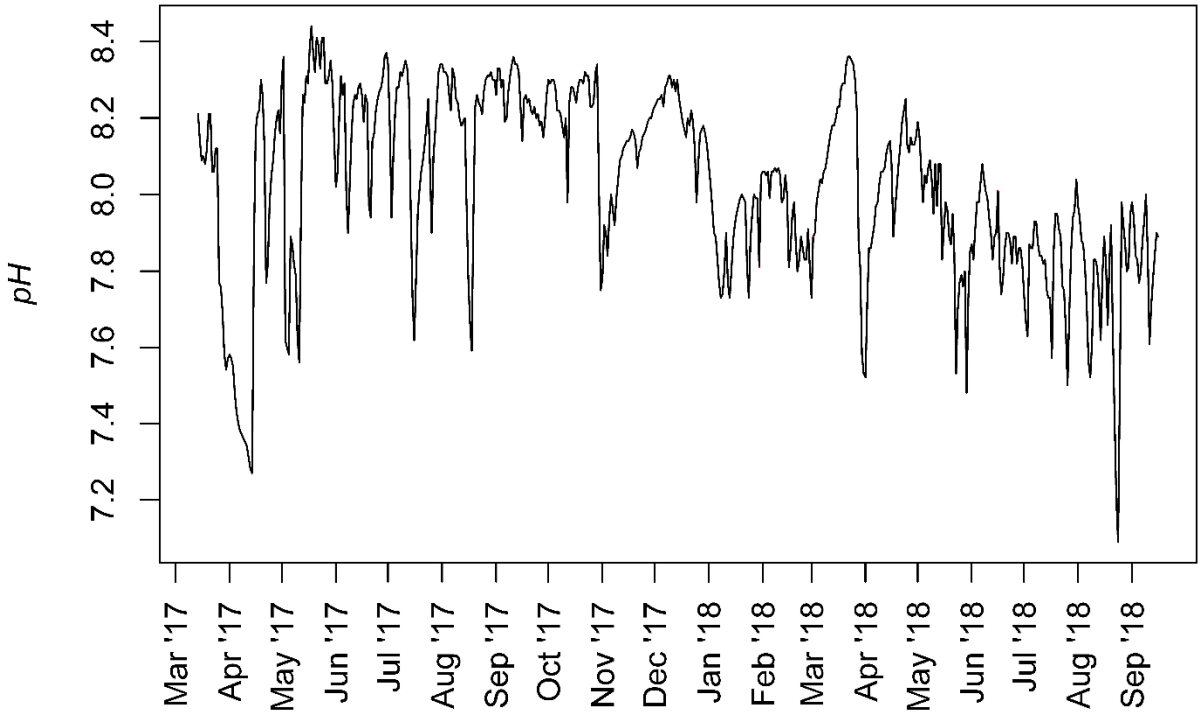



**Figure C33:** Temporal profile of average daily pH (*pH*).

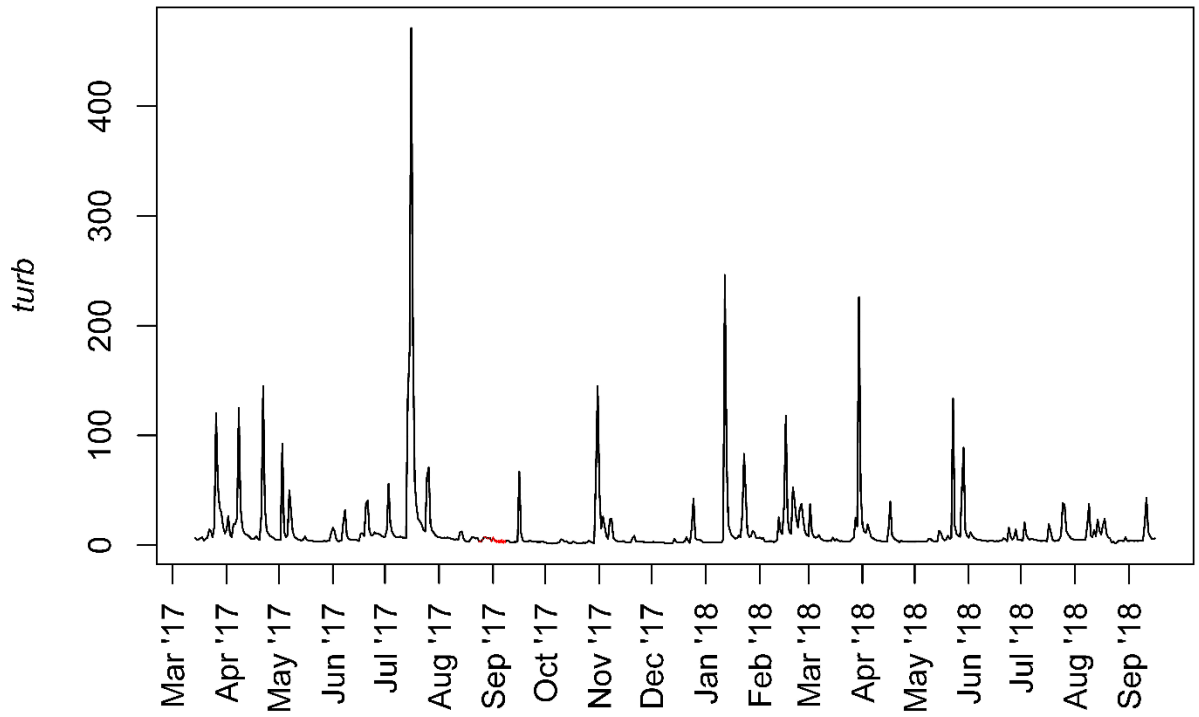



**Figure C34:** Temporal profile of average daily turbidity (*turb*) in NTU.

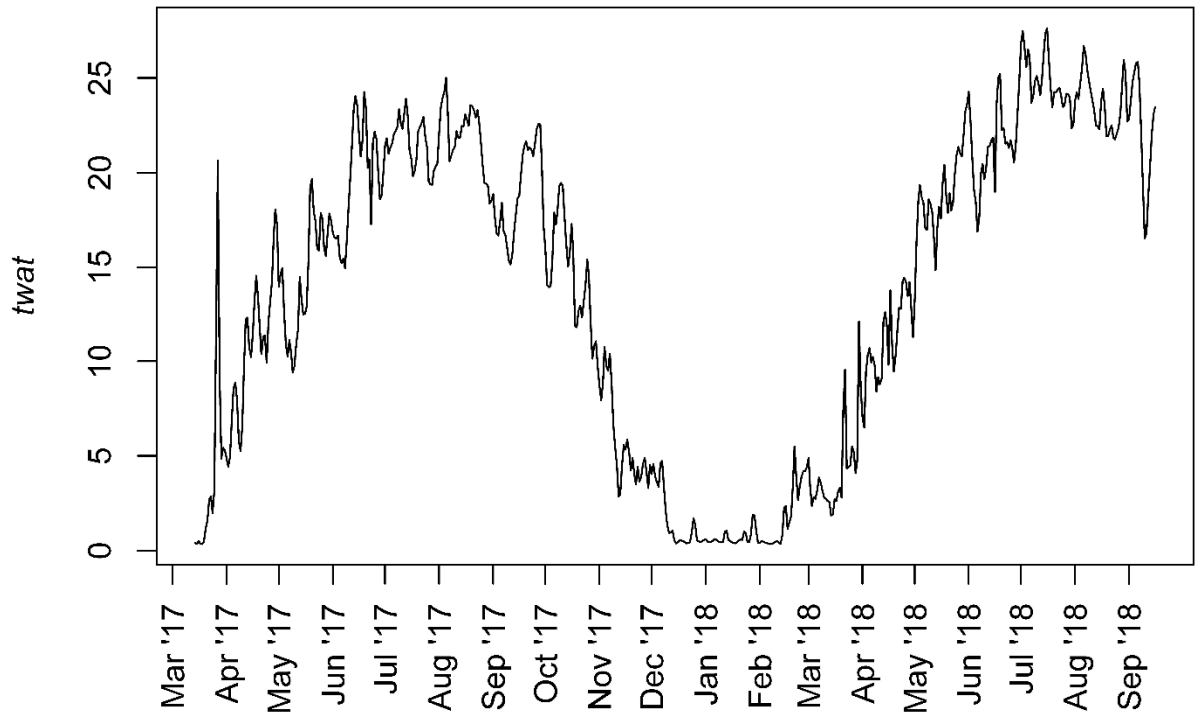



**Figure C35:** Temporal profile of average daily water temperature (*twat*) in °C.

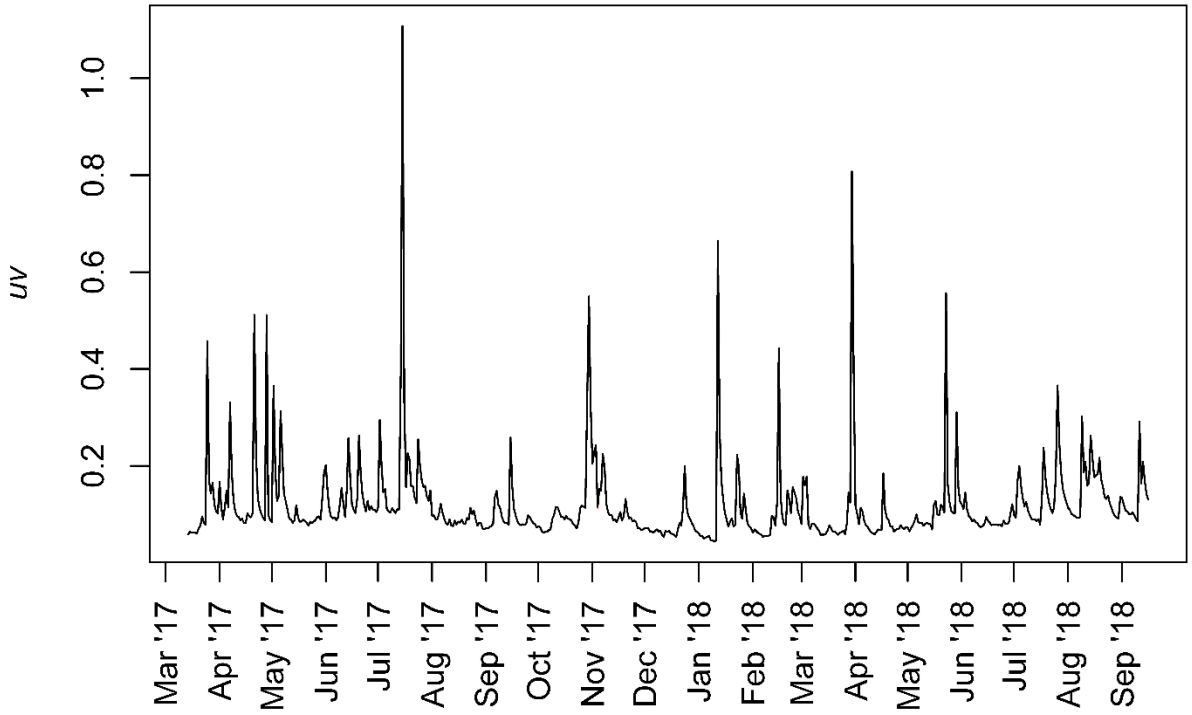



**Figure C36:** Temporal profile of average daily $UV_{254}$ absorbance (*uv*).

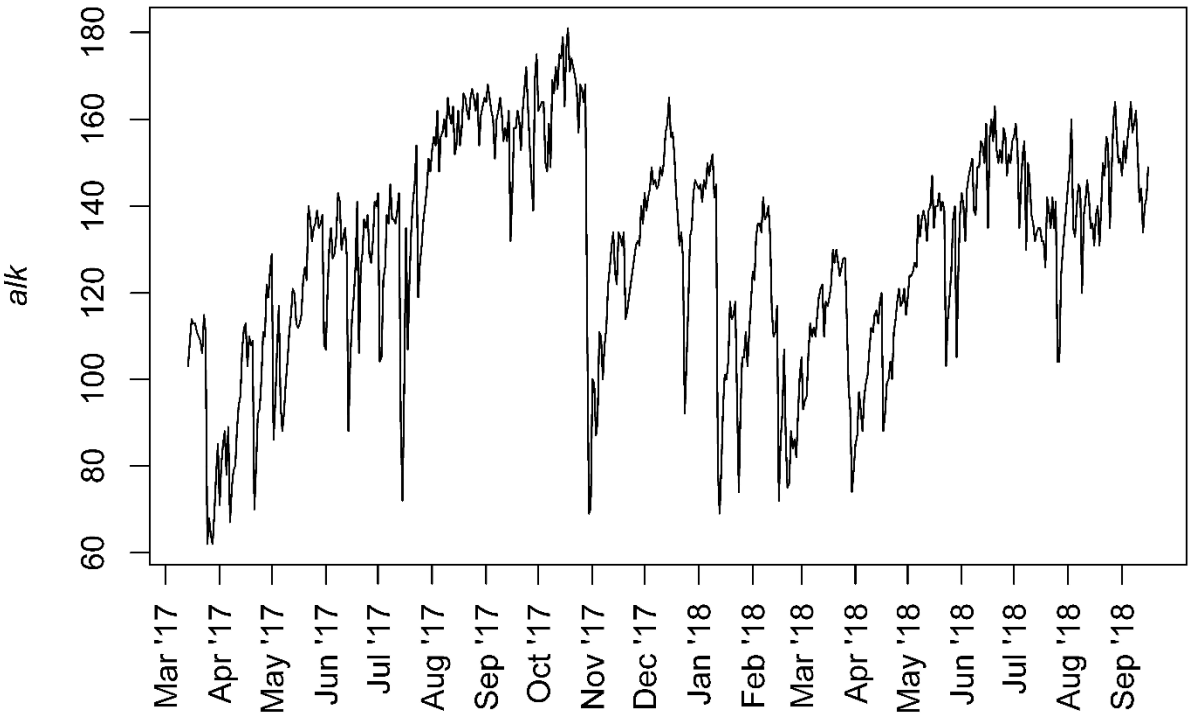



**Figure C37:** Temporal profile of daily alkalinity measurements (*alk*) in $mg \cdot L^{-1}$ as $CaCO_3$.

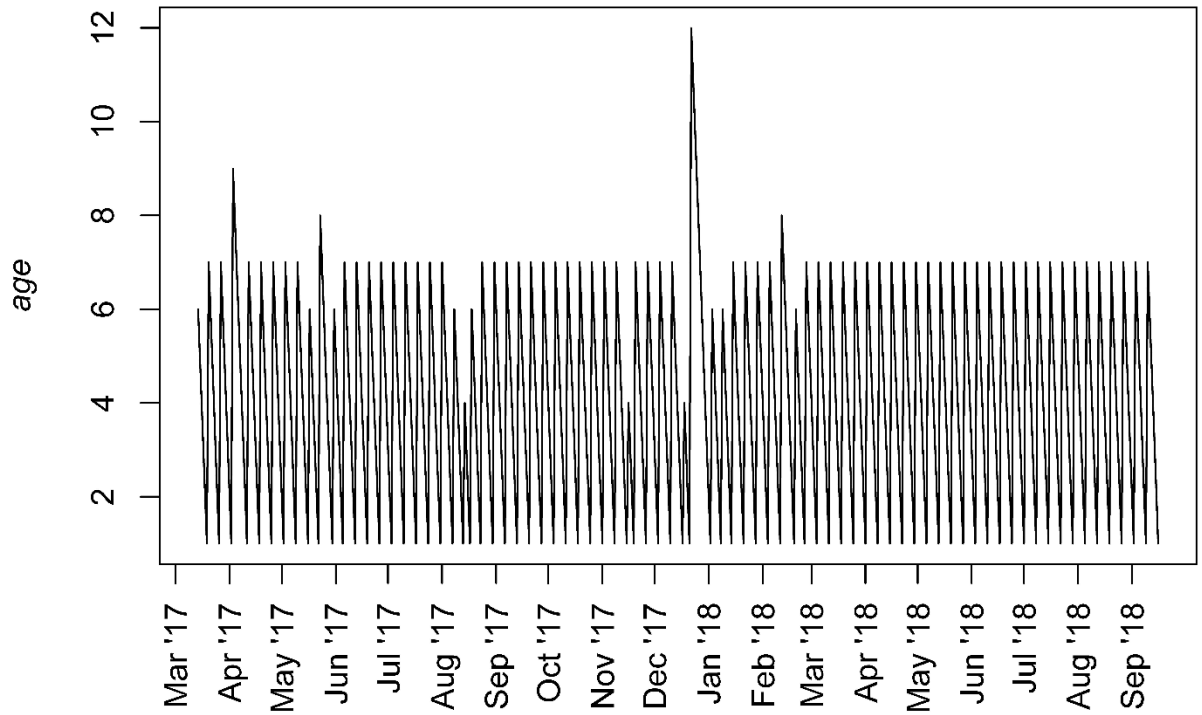



**Figure C38:** Temporal profile of average daily water sample age (*age*) in days.

**Figure C39** describes the optimization of the apparent growing seasons for ATR. For each growing season (2017 and 2018), the cross-correlation coefficients between ATR concentrations and *bfi* were examined along different assumed growing seasons that varied by start date and duration. Each point on the figure describes the correlation coefficient for an assumed growing season. Optimal apparent growing seasons were selected by maximizing the correlation coefficients and durations. Based on our results, *bfi* is only well-correlated with the agriculture-derived micropollutants during their respective application periods.

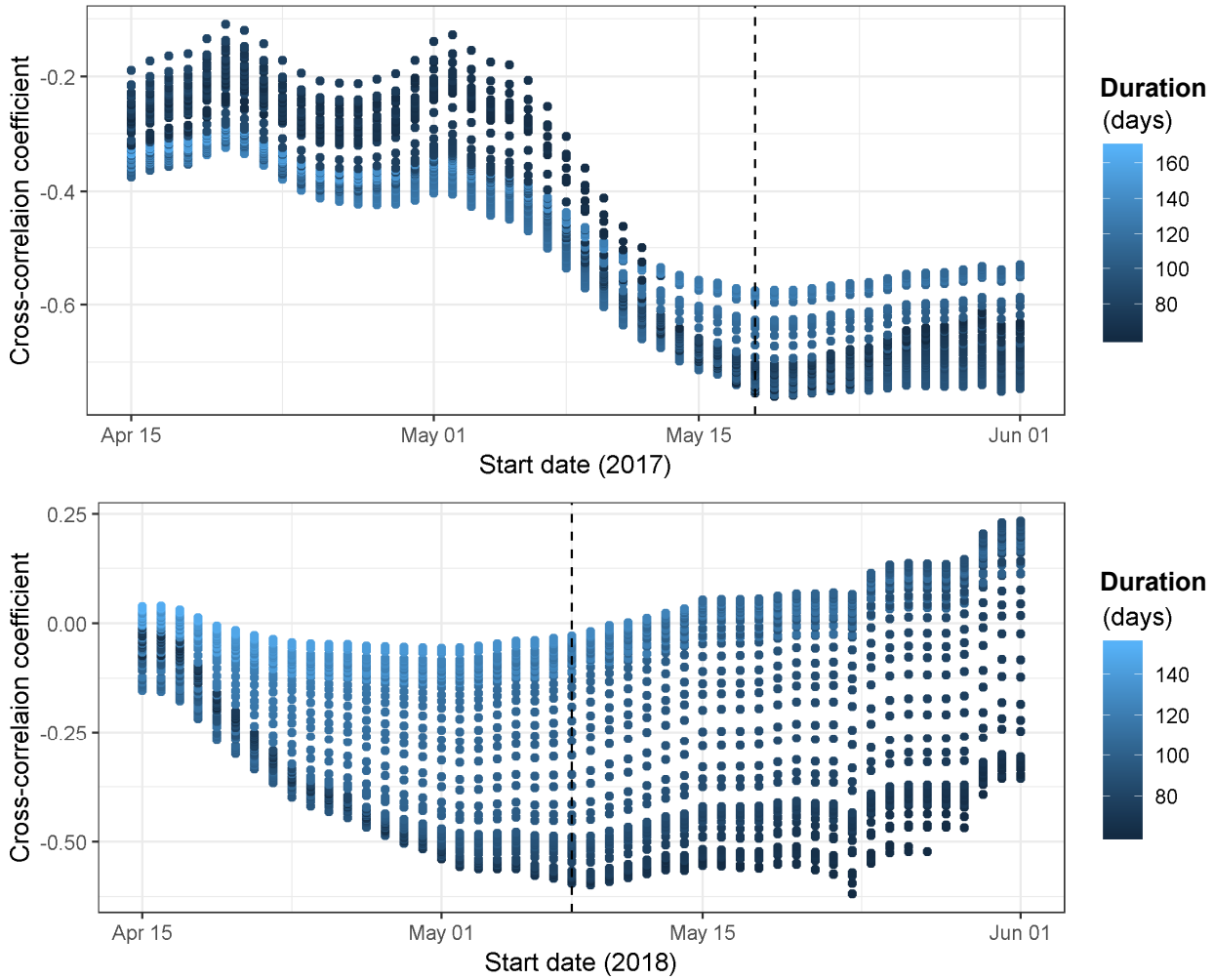



**Figure C39:** Optimal growing seasons for atrazine as determined by correlation with *bfi*. Optimal start dates are marked with a dashed line.

## C.3 Multivariable regression

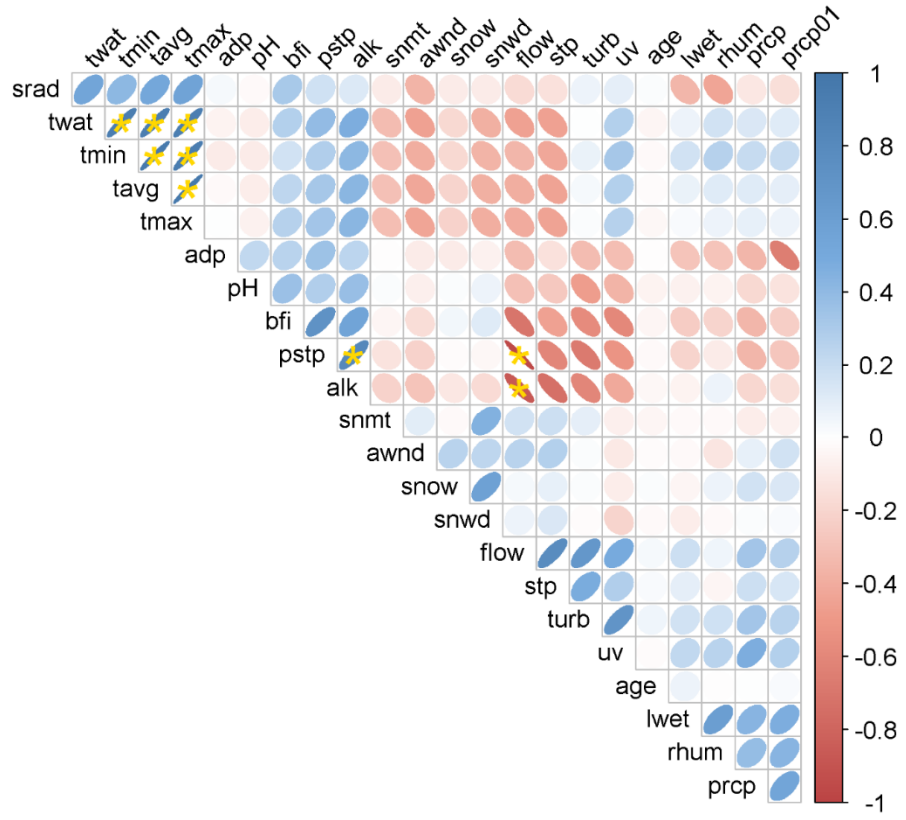



**Figure C40:** Environmental covariates (average of the imputed datasets) correlation matrix. Significant Pearson correlations ($\rho > 0.8$, p-value $< 0.01$) are marked (*).

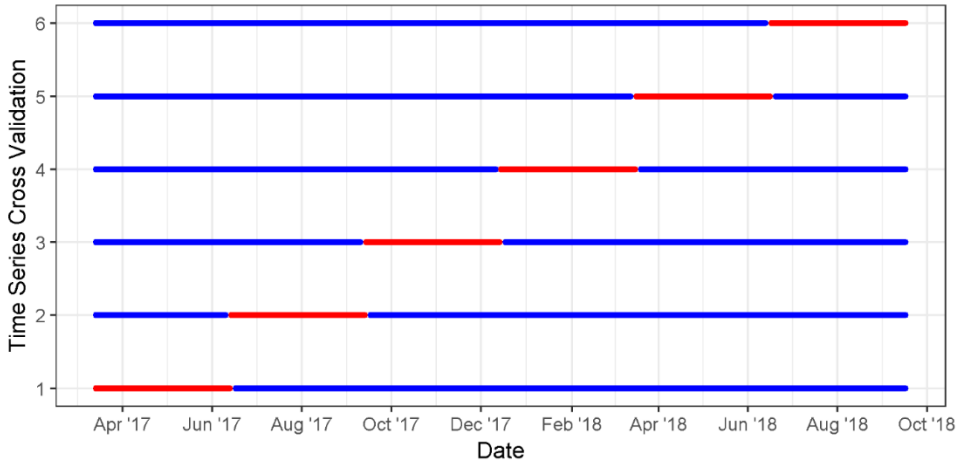



**Figure C41:** Cross-validation time series; blue lines represent training sets and red lines represent test sets.

**Figures C42 – C61** describe the bivariate regression plots and significant final standardized coefficients for the generalized least squares (GLS) and logistic (logit) regressions following the model framework outlined in Chapter 4. The bivariate regression plots include all environmental covariates in the best subsets regression model equations for each micropollutant profile. The interaction terms between *grow* and the environmental covariates are specified as *grow_X* and *non_grow_X*. Each imputed dataset is plotted at 20% opacity (non-imputed points are solid, while imputed points are semitransparent) and best fit lines (red) for each imputed dataset are shown. Logit regression plots show the density histograms for peak events (top, red) and non-peak events (bottom, blue). The final standardized coefficient plots represent the best combination of environmental covariates based on corrected Akaike information criterion (AICc) and multimodel inference. The boxplots represent the coefficients determined for each cross validation sample set for each imputed dataset (n = 30), and the blue points represent the overall mean coefficient value (error bars represent the 95% confidence interval). Train and test set model diagnostics (GLS, adjusted $R^2$; logit, accuracy) are provided for each micropollutant profile.

## Atrazine (ATR)

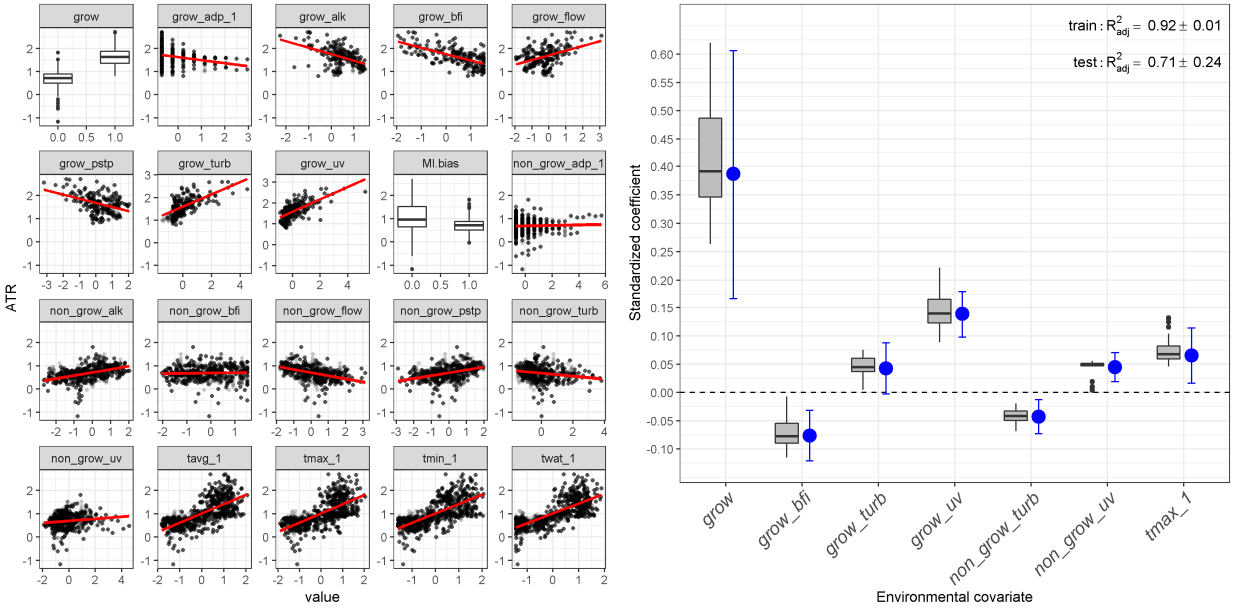



**Figure C42:** Bivariate GLS regression plots for ATR (left) and significant final standardized coefficients (right).

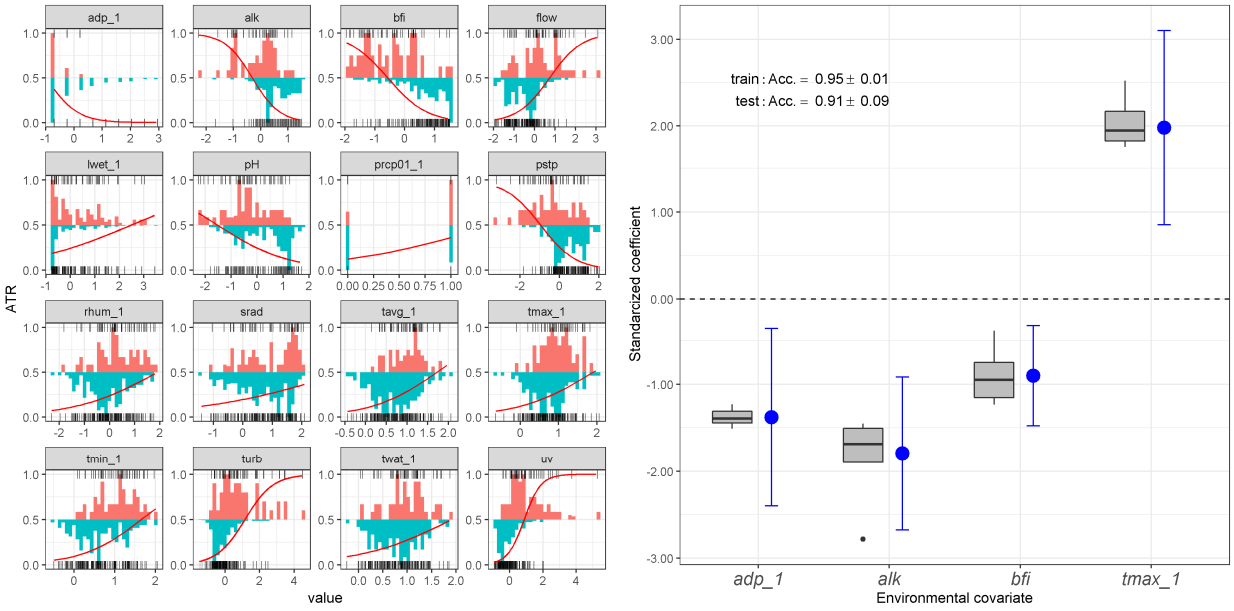



**Figure C43:** Bivariate logit regression plots for ATR (left) and significant final standardized coefficients (right) during the growing season.

## Atrazine-desethyl (ATR.d)

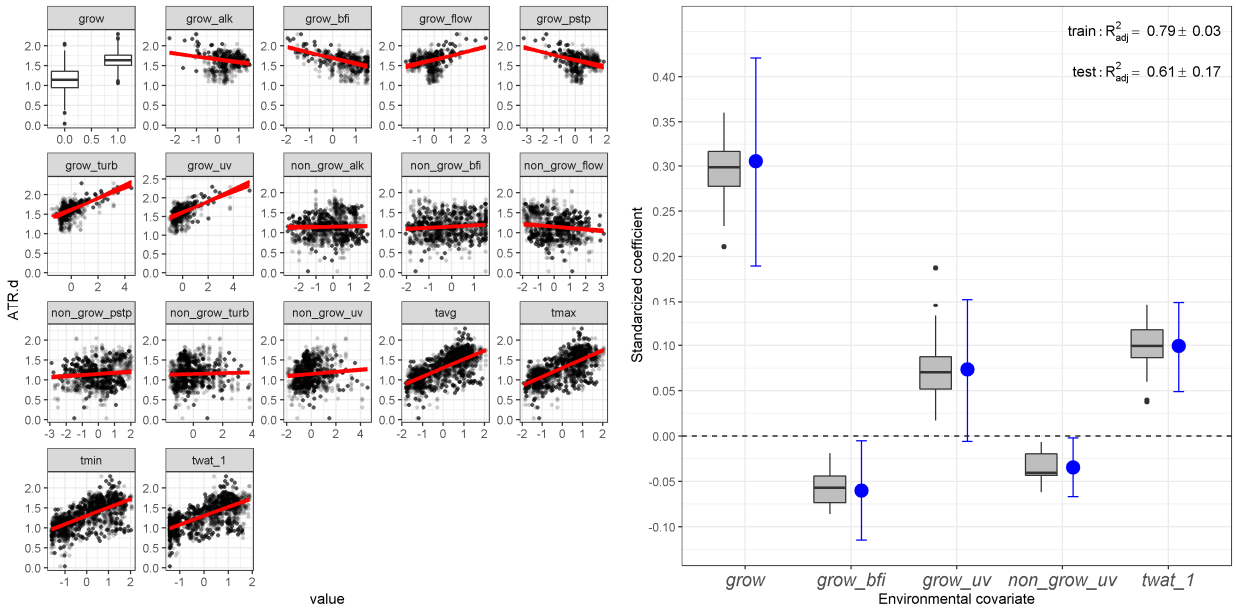



**Figure C44:** Bivariate GLS regression plots for ATR.d (left) and significant final standardized coefficients (right).

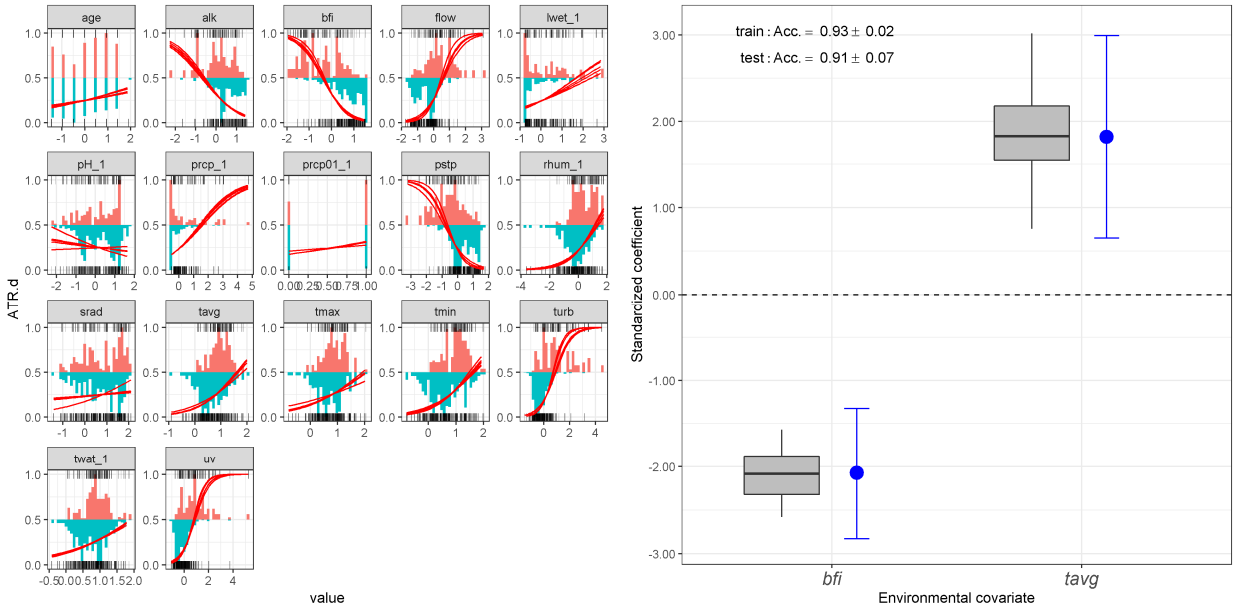



**Figure C45:** Bivariate logit regression plots for ATR.d (left) and significant final standardized coefficients (right) during the growing season.

## Atrazine-2-hydroxy (ATR.h)

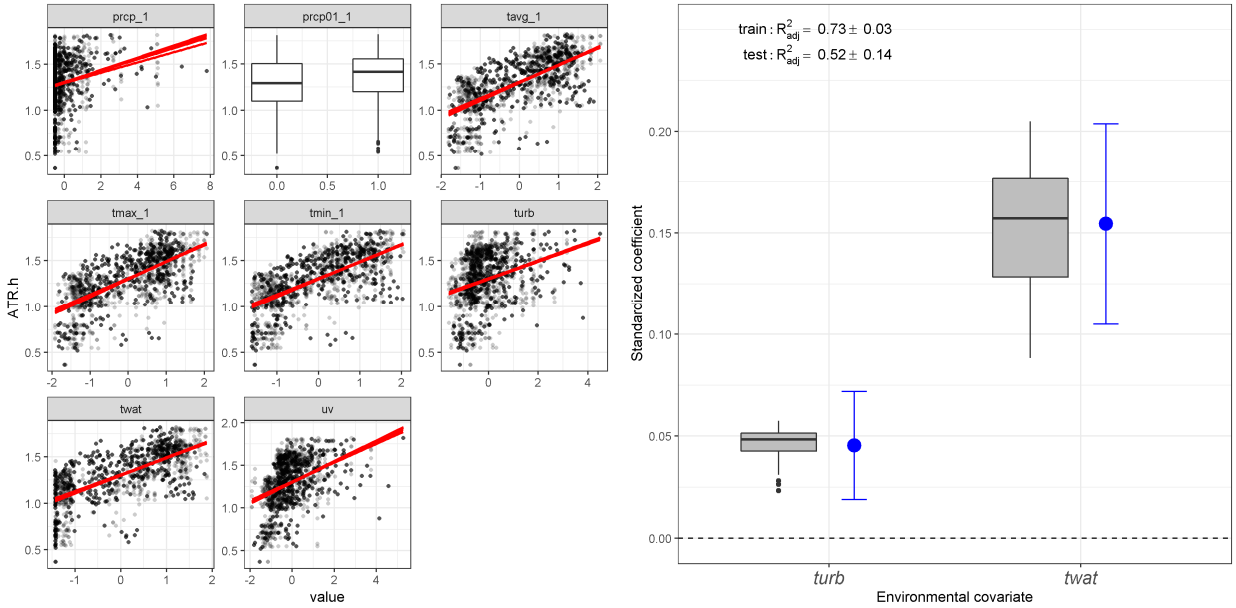



**Figure C46:** Bivariate GLS regression plots for ATR.h (left) and significant final standardized coefficients (right).

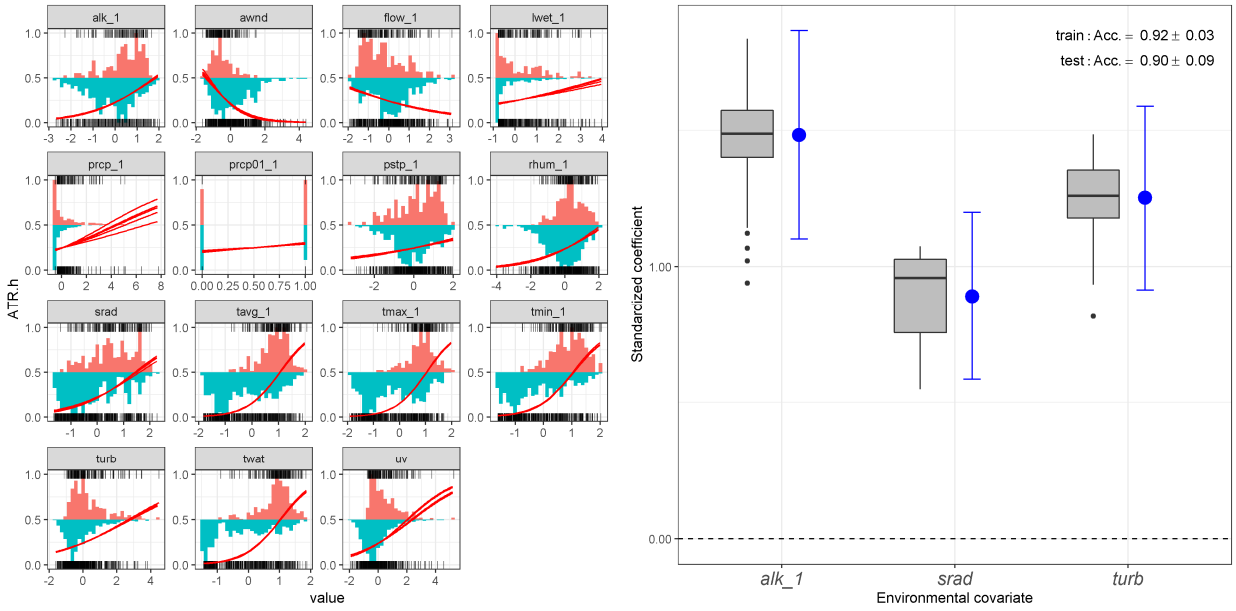



**Figure C47:** Bivariate logit regression plots for ATR.h (left) and significant final standardized coefficients (right).

## Metolachlor (MET)

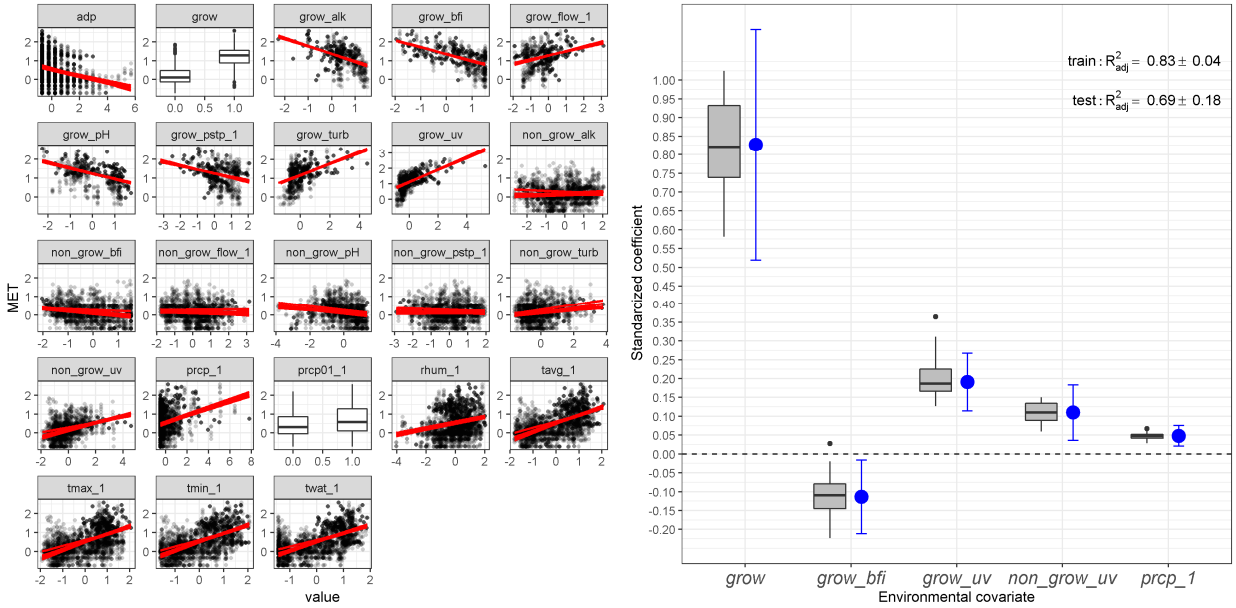



**Figure C48:** Bivariate GLS regression plots for MET (left) and significant final standardized coefficients (right).

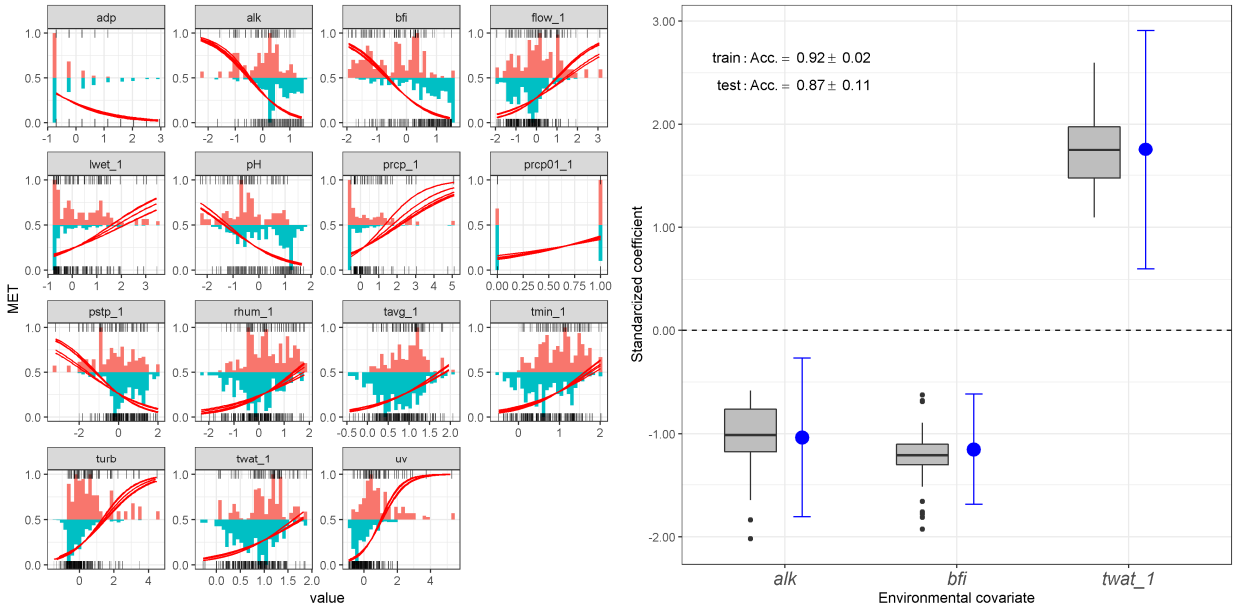



**Figure C49:** Bivariate logit regression plots for MET (left) and significant final standardized coefficients (right) during the growing season.

## Metolachlor-OXA (MET.o)

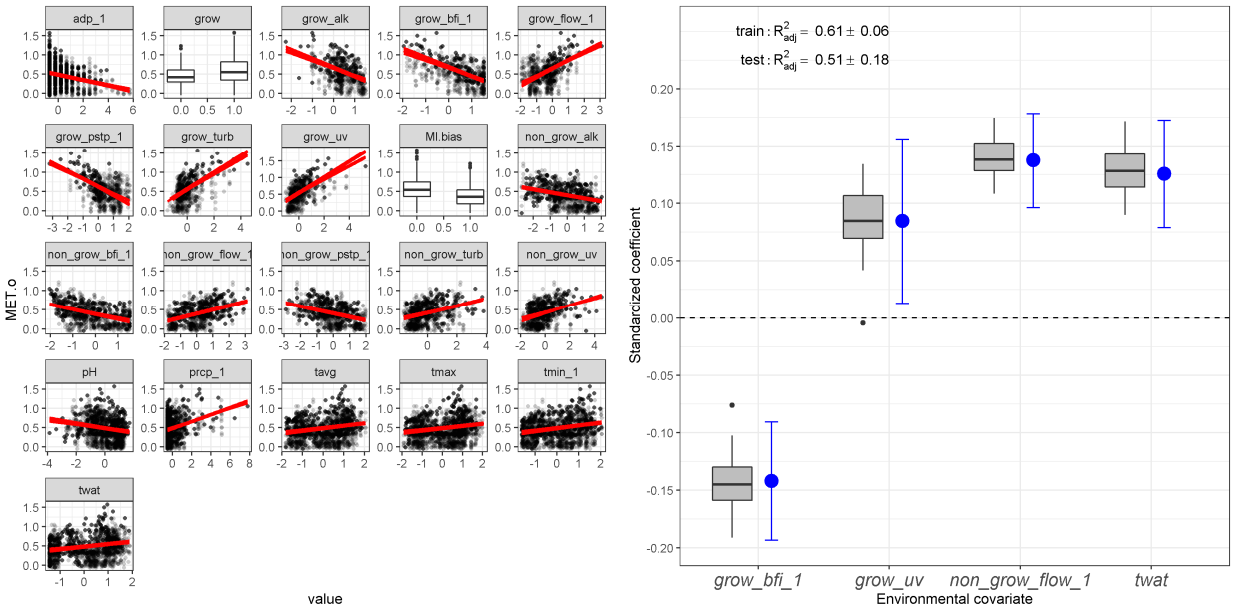



**Figure C50:** Bivariate GLS regression plots for MET.o (left) and significant final standardized coefficients (right).

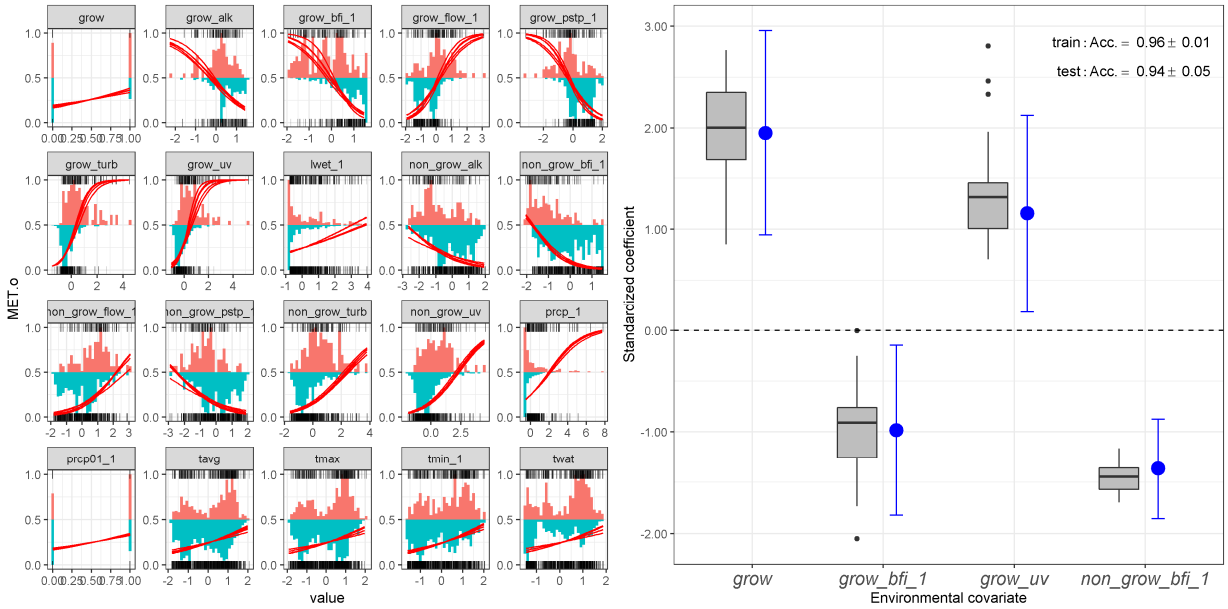



**Figure C51:** Bivariate logit regression plots for MET.o (left) and significant final standardized coefficients (right).

## Desvenlafaxine (DES)

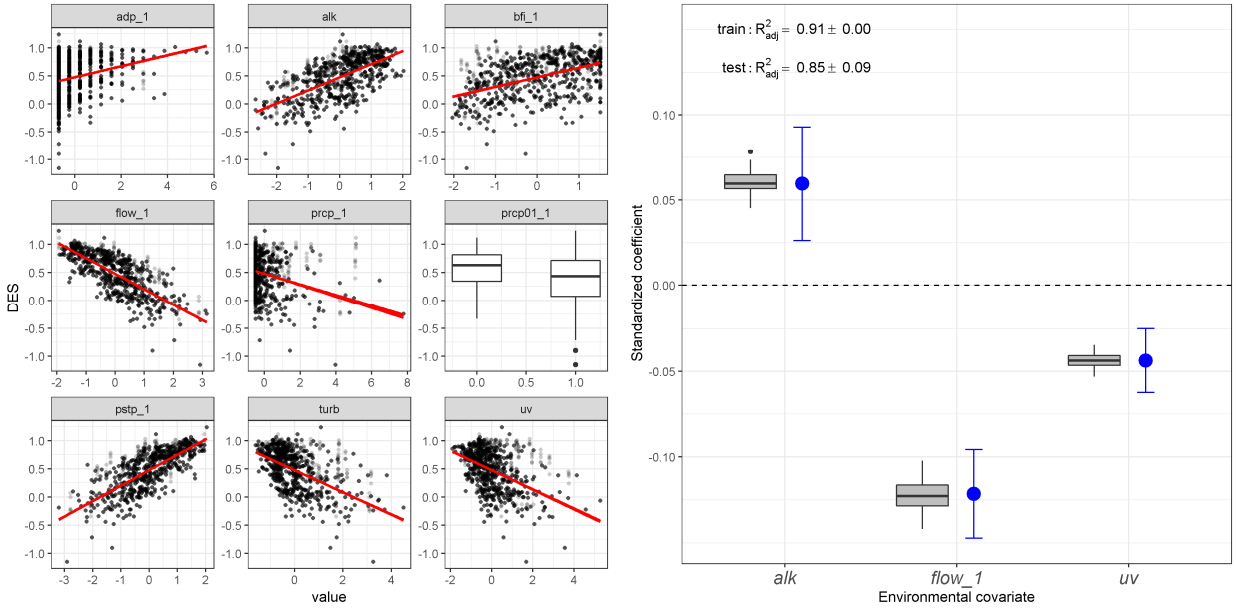



**Figure C52:** Bivariate GLS regression plots for DES (left) and significant final standardized coefficients (right).

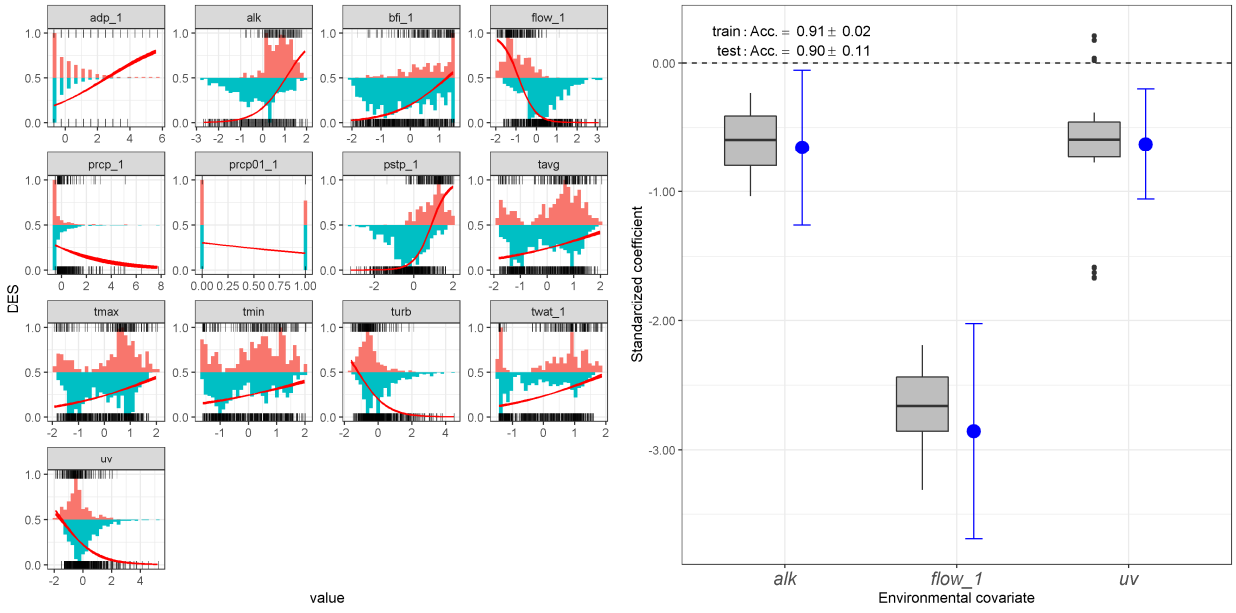



**Figure C53:** Bivariate logit regression plots for DES (left) and significant final standardized coefficients (right).

## Rac-threo-dihydrobupropion (DHB)

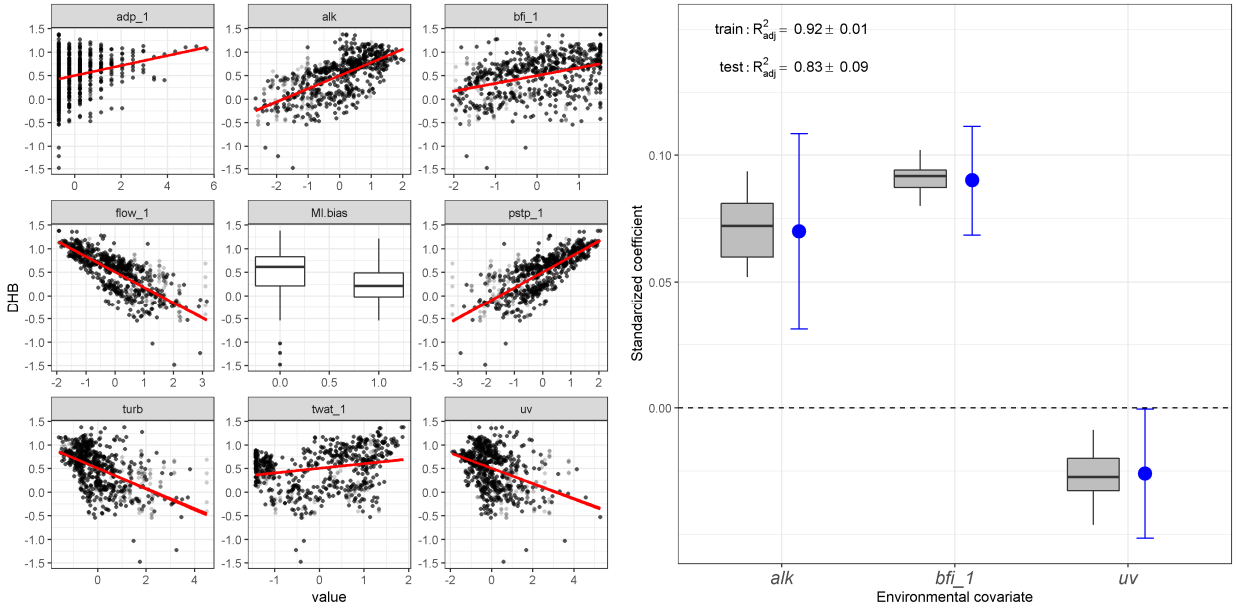



**Figure C54:** Bivariate GLS regression plots for DHB (left) and significant final standardized coefficients (right).

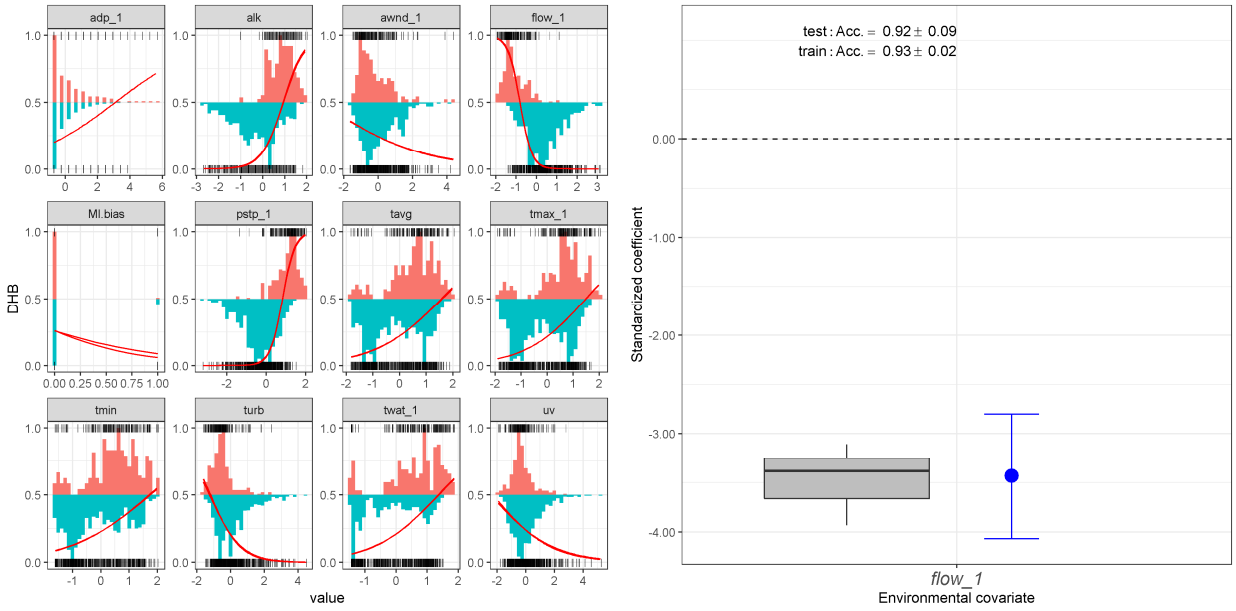



**Figure C55:** Bivariate logit regression plots for DHB (left) and significant final standardized coefficients (right).

## Gabapentin-lactam (GAB.l)

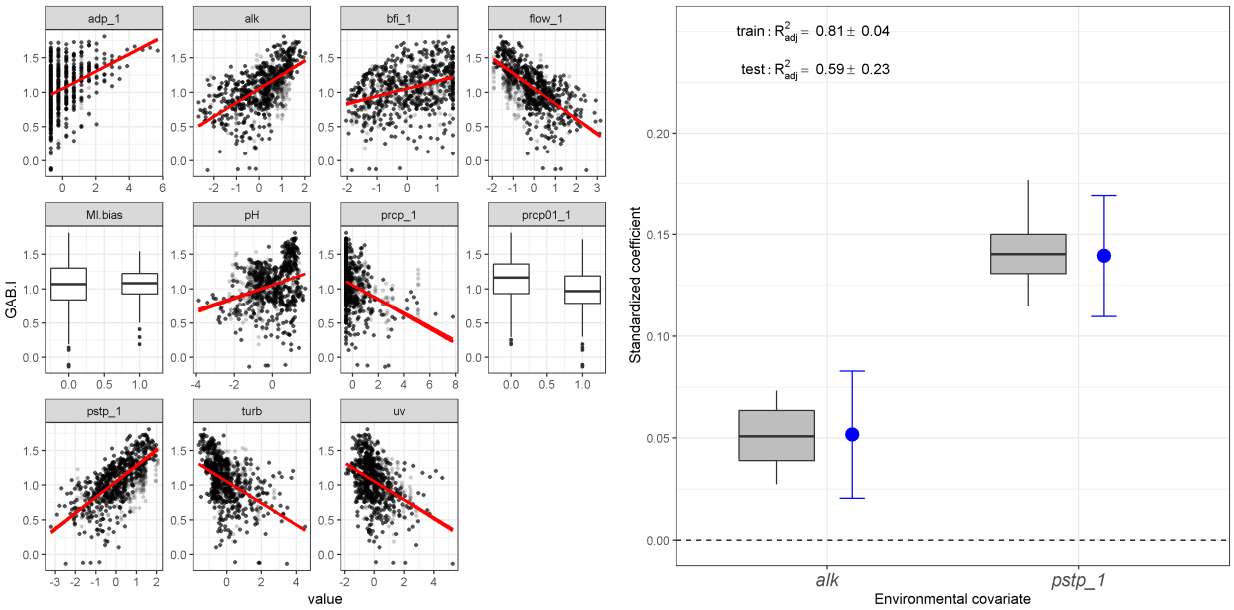



**Figure C56:** Bivariate GLS regression plots for GAB.l (left) and significant final standardized coefficients (right).

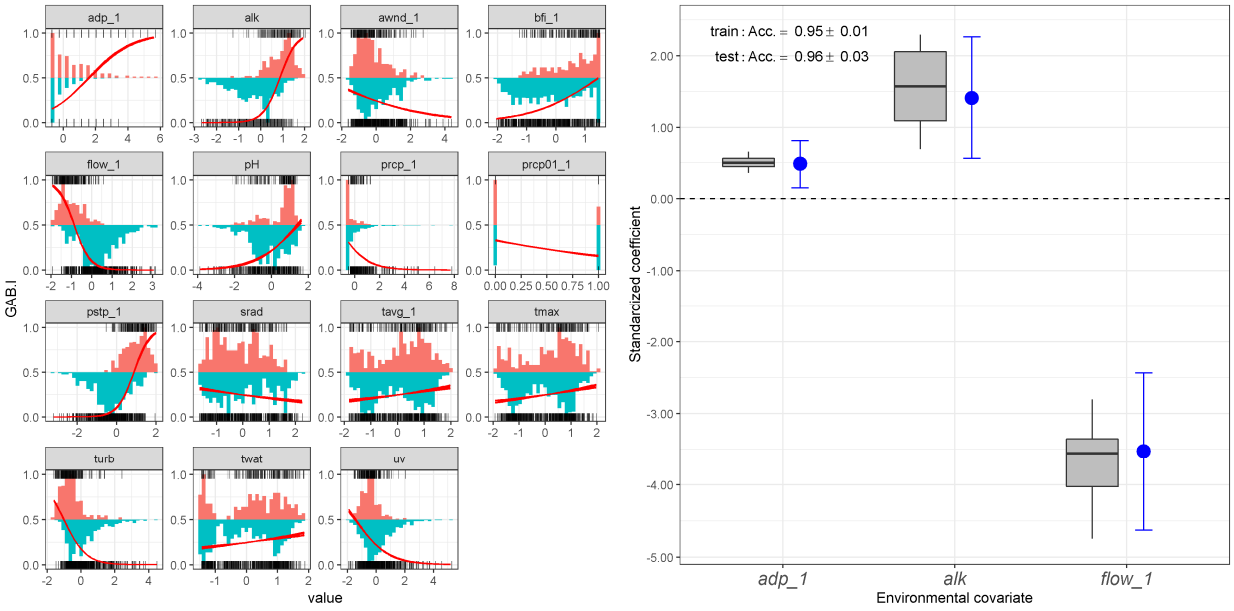



**Figure C57:** Bivariate logit regression plots for GAB.l (left) and significant final standardized coefficients (right).

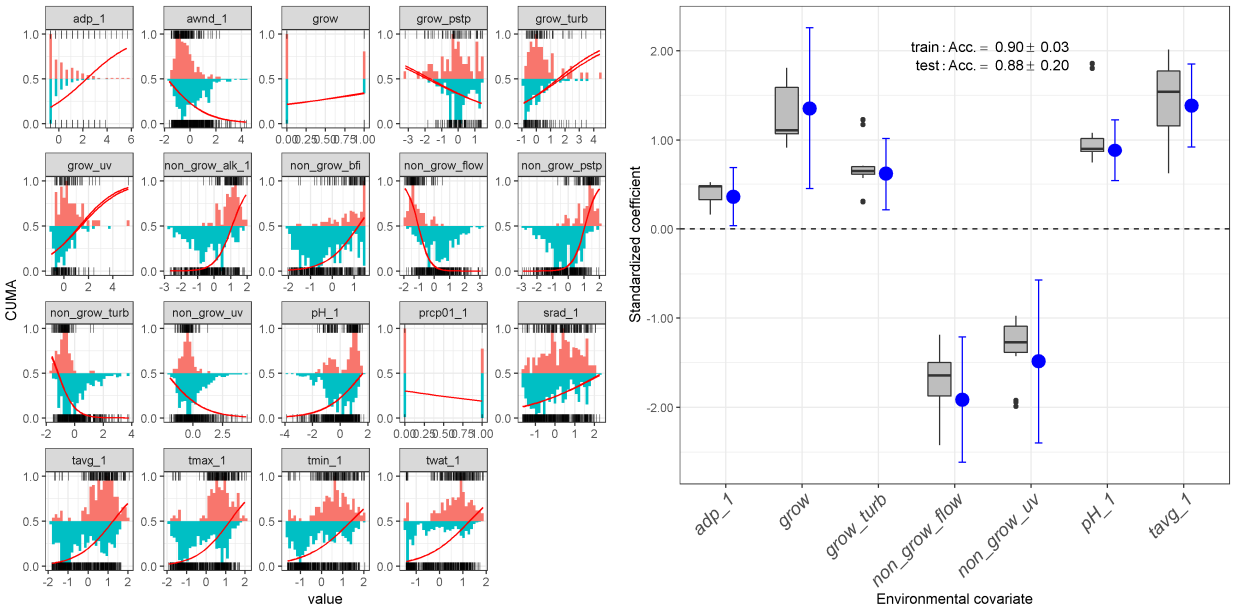



**Figure C58:** Bivariate logit regression plots for CUMA (left) and significant final standardized coefficients (right).

## Cumulative estimated concentration (CUMC)

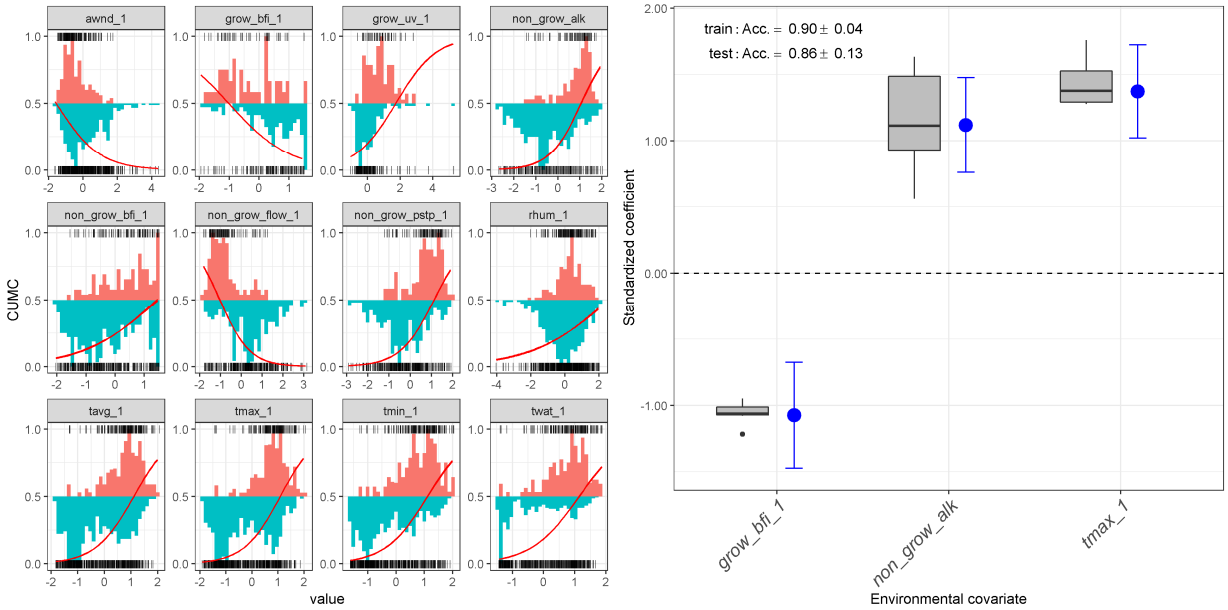



**Figure C59:** Bivariate logit regression plots for CUMC (left) and significant final standardized coefficients (right).

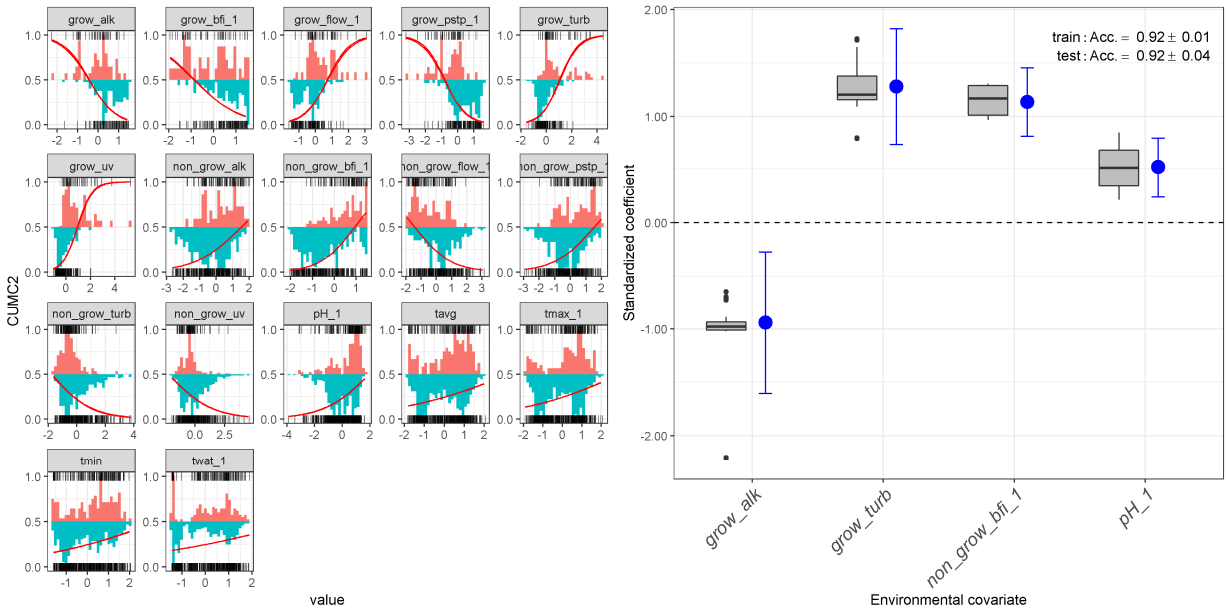



**Figure C60:** Bivariate logit regression plots for CUMC2 (left) and significant final standardized coefficients (right).

## Cumulative estimated toxicity (CUMT)

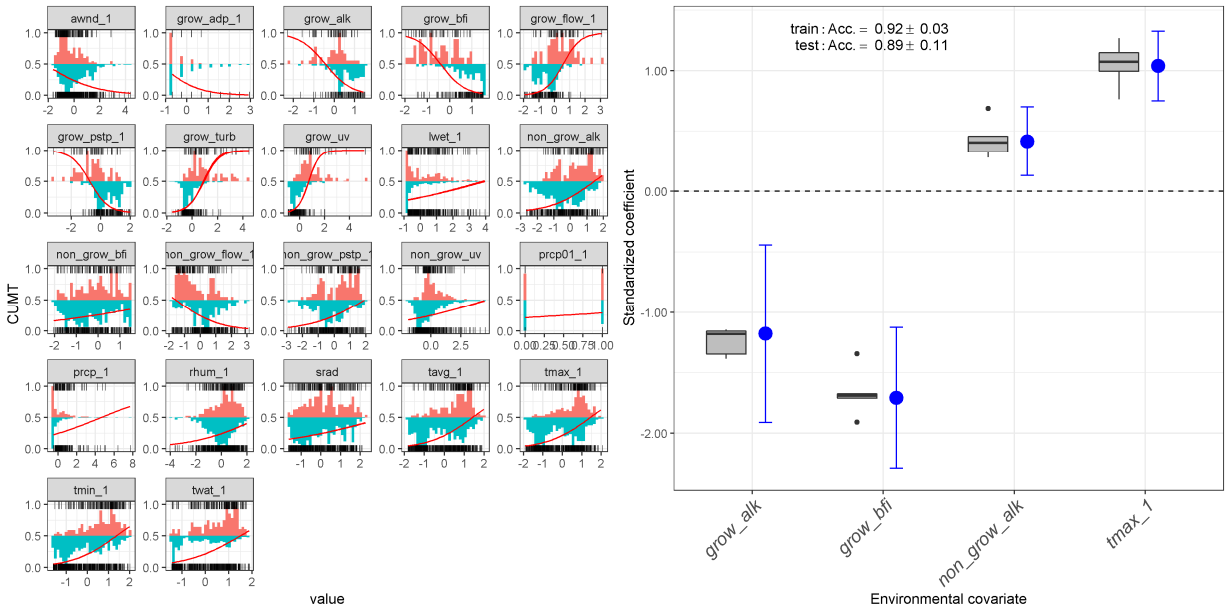
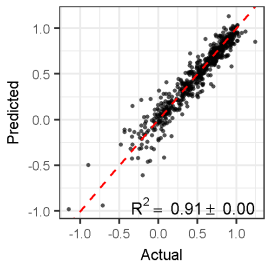
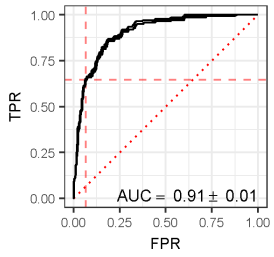
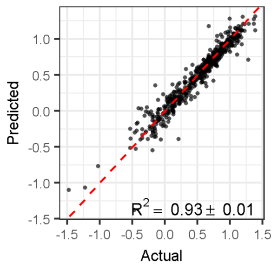
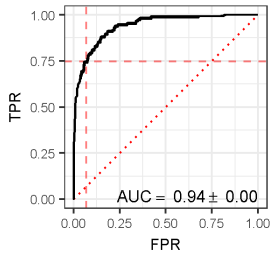
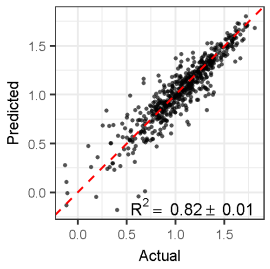
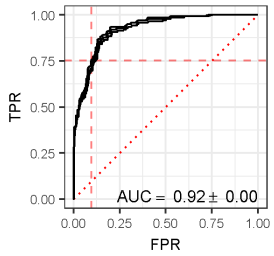
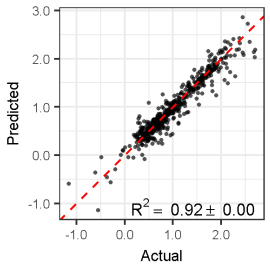
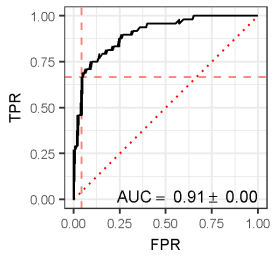
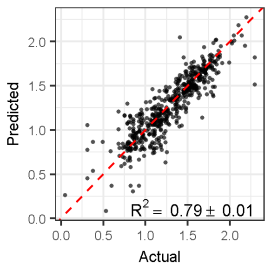
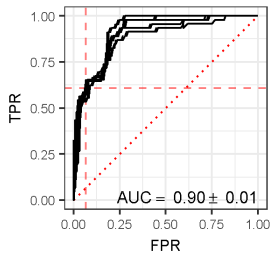
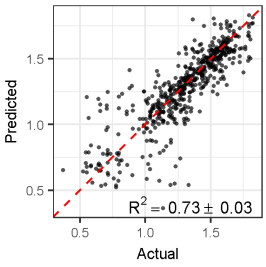
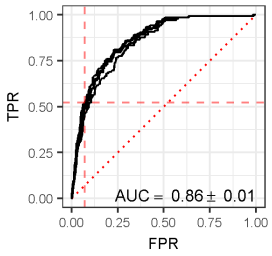
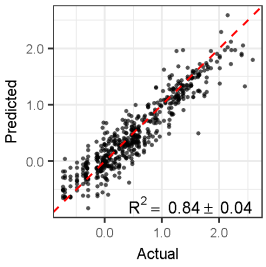
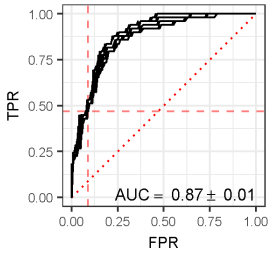
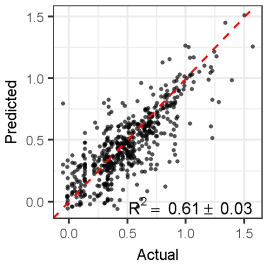
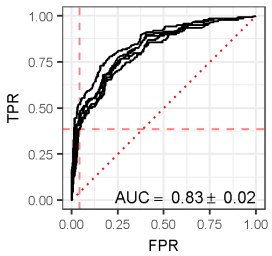



**Figure C61:** Bivariate logit regression plots for CUMT (left) and significant final standardized coefficients (right).

**Table C1:** Regression results using the final averaged coefficients for micropollutant dynamics and peak events of representative micropollutant profiles. Significant covariates are listed for each micropollutant profile with the sign of the relationship (+/–). Final model diagnostics ($R^2$ and area under curve, AUC) represent each imputed dataset (average ± s.d.).

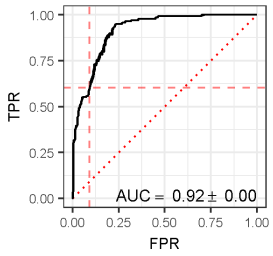
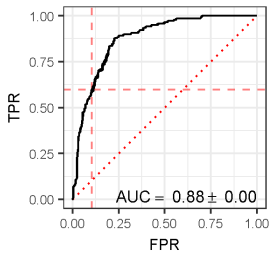
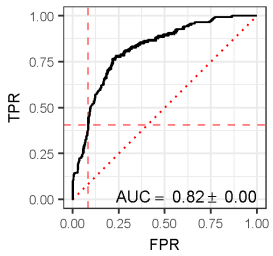
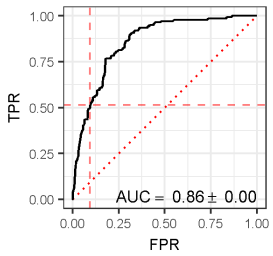
| Profile | Micropollutant Dynamics (GLS regression)[a] | | Predicting Peak Events (logit regression)[b] | |
|---|---|---|---|---|
| **DES** | + *alk*<br>– $flow_{t-1}$<br>– *uv* |  | + *alk*<br>– $flow_{t-1}$<br>– *uv* |  |
| **DHB** | + *alk*<br>+ $bfi_{t-1}$<br>– *uv* |  | – $flow_{t-1}$ |  |
| **GAB.l** | + *alk*<br>+ $pstp_{t-1}$ |  | – $flow_{t-1}$<br>+ $adp_{t-1}$<br>+ *alk* |  |
| **ATR** | + *grow*<br>– *grow_bfi*<br>+ *grow_turb*<br>+ *grow_uv*<br>– *nongrow_turb*<br>+ *nongrow_uv*<br>+ $tmax_{t-1}$ |  | – $grow\_adp_{t-1}$<br>– *grow_alk*<br>– *grow_bfi*<br>+ $grow\_tmax_{t-1}$ |  |
| **ATR.d** | + *grow*<br>– *grow_bfi*<br>+ *grow_uv*<br>+ *nongrow_uv*<br>+ $twat_{t-1}$ |  | – *grow_bfi*<br>+ *grow_tavg* |  |

| Profile | Micropollutant Dynamics (GLS regression)[a] | Predicting Peak Events (logit regression)[b] |
|---|---|---|
| **ATR.h** | + *turb*<br>+ *twat*<br>  | + $alk_{t-1}$<br>+ *srad*<br>+ *turb*<br>  |
| **MET** | + *grow*<br>– *grow_bfi*<br>+ *grow_uv*<br>+ *nongrow_uv*<br>+ $prcp_{t-1}$<br>  | – *grow_alk*<br>– *grow_bfi*<br>+ $grow\_twat_{t-1}$<br>  |
| **MET.o** | – $grow\_bfi_{t-1}$<br>+ *grow_uv*<br>+ $nongrow\_flow_{t-1}$<br>+ *twat*<br>  | + *grow*<br>– $grow\_bfi_{t-1}$<br>+ *grow_uv*<br>– $nongrow\_bfi_{t-1}$<br>  |

[a] GLS regression diagnostic plots represent the actual profile values vs. predicted values using final regression coefficients; dashed line = 1:1.

[b] Logit regression diagnostic plots represent receiver operating characteristic (ROC) curves; FPR = false positive rate (1 – specificity); TPR = true positive rate (sensitivity); dashed line = cutoff at probability of 0.5; dotted line = random guess, 1:1.

**Table C2:** Regression results using the final averaged coefficients for peak events of cumulative profiles. Significant covariates are listed for each micropollutant profile with the sign of the relationship (+/–). Final model diagnostics (area under curve, AUC) represent each imputed dataset (average ± s.d.).

| Profile | Predicting Peak Events (logit regression)[a] | |
|---|---|---|
| **CUMA** | + $adp_{t-1}$<br>+ *grow*<br>+ *grow_turb*<br>– *nongrow_flow*<br>– *nongrow_uv*<br>+ $pH_{t-1}$<br>+ $tavg_{t-1}$ |  |
| **CUMC** | – $grow\_bfi_{t-1}$<br>+ *nongrow_alk*<br>+ $tmax_{t-1}$ |  |
| **CUMC2** | – *grow_alk*<br>+ *grow_turb*<br>+ $nongrow\_bfi_{t-1}$<br>+ $pH_{t-1}$ |  |
| **CUMT** | – *grow_alk*<br>– *grow_bfi*<br>+ *nongrow_alk*<br>+ $tmax_{t-1}$ |  |

[a] Logit regression diagnostic plots represent receiver operating characteristic (ROC) curves; FPR = false positive rate (1 – specificity); TPR = true positive rate (sensitivity); dashed line = cutoff at probability of 0.5; dotted line = random guess, 1:1.

## C.4 Recommendations for future micropollutant sampling strategies

**Table C3:** Percent of environmental risk captured using uninformed and informed micropollutant sampling strategies. Informed strategies included the final peak event model and important bivariate models. The percent of environmental risk captured by each sampling strategy was determined by comparing the areas under the resulting profiles with the areas under the actual representation of risk profiles.

| **Profile** | | ***Uninformed***[a] | ***Informed*** | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | ***final*** | ***bfi*** | ***flow*** | ***alk*** | ***turb*** | ***uv*** | ***prcp*** |
| **ATR** | *weekly* | 53 ± 10% | 99% | 92% | 92% | 98% | 75% | 95% | 76% |
| | *monthly* | 29 ± 8% | 106% | 92% | 92% | 84% | 72% | 100% | 40% |
| **DES** | *weekly* | 82 ± 3% | 87% | 88% | 84% | 86% | 87% | 86% | n.a. |
| | *monthly* | 61 ± 4% | 75% | 77% | 82% | 73% | 74% | 72% | n.a. |
| **CUMC** | *weekly* | 81 ± 3% | 91% | 84% | 84% | 84% | 82% | 83% | 84% |
| | *monthly* | 64 ± 4% | 70% | 73% | 69% | 66% | 67% | 69% | 70% |
| **CUMC2** | *weekly* | 69 ± 4% | 83% | 81% | 83% | 86% | 77% | 81% | 73% |
| | *monthly* | 45 ± 4% | 65% | 67% | 69% | 65% | 63% | 77% | 48% |
| **CUMT** | *weekly* | 66 ± 4% | 92% | 84% | 84% | 91% | 75% | 84% | 79% |
| | *monthly* | 45 ± 6% | 85% | 79% | 78% | 76% | 67% | 93% | 54% |

[a]Values represent the mean ± standard deviation of area ratios for each uninformed starting date ($n_{weekly}$ = 7; $n_{monthly}$ = 30).

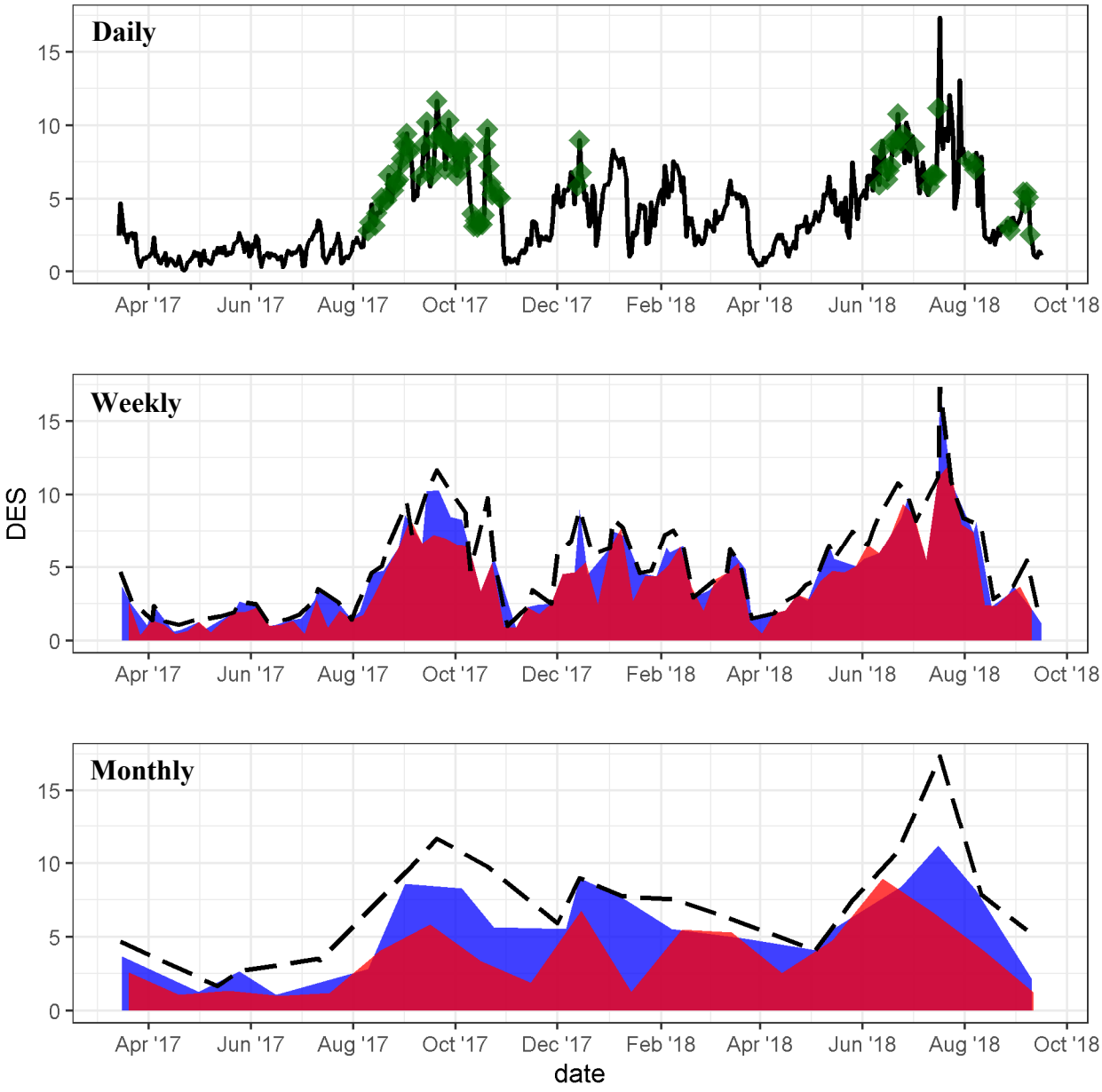



**Figure C62:** Informed vs uninformed sampling strategies for desvenlafaxine (DES) concentration in ng·L$^{-1}$; solid black line = average of the imputed datasets; green points = triggered sampling events (≥2); dashed black line = actual representation of risk; blue area = informed sampling, red area = uninformed sampling.

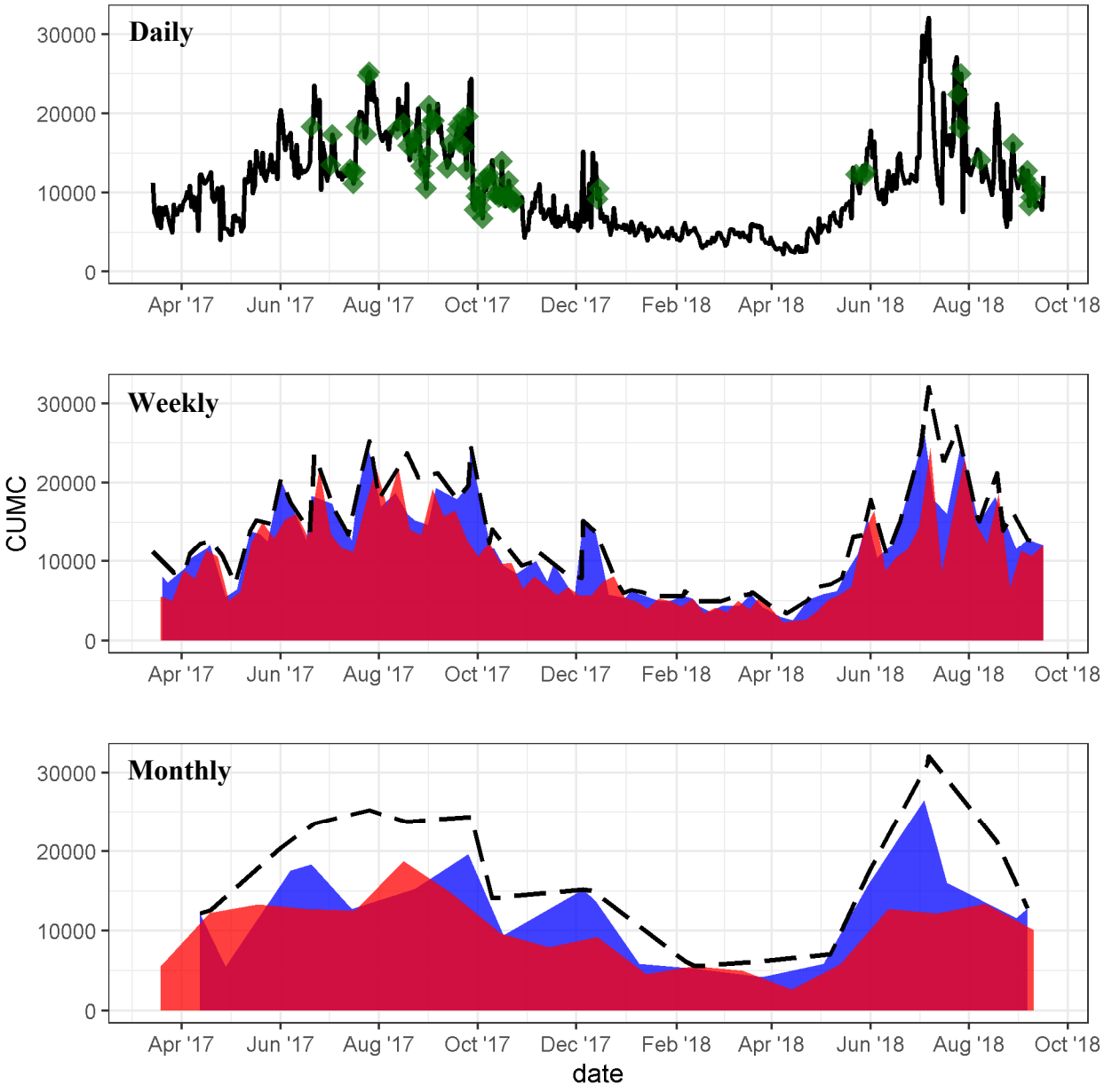



**Figure C63:** Informed vs uninformed sampling strategies for cumulative estimated concentration (CUMC) in ng·L$^{-1}$; solid black line = average of the imputed datasets; green points = triggered sampling events (≥2); dashed black line = actual representation of risk; blue area = informed sampling, red area = uninformed sampling.

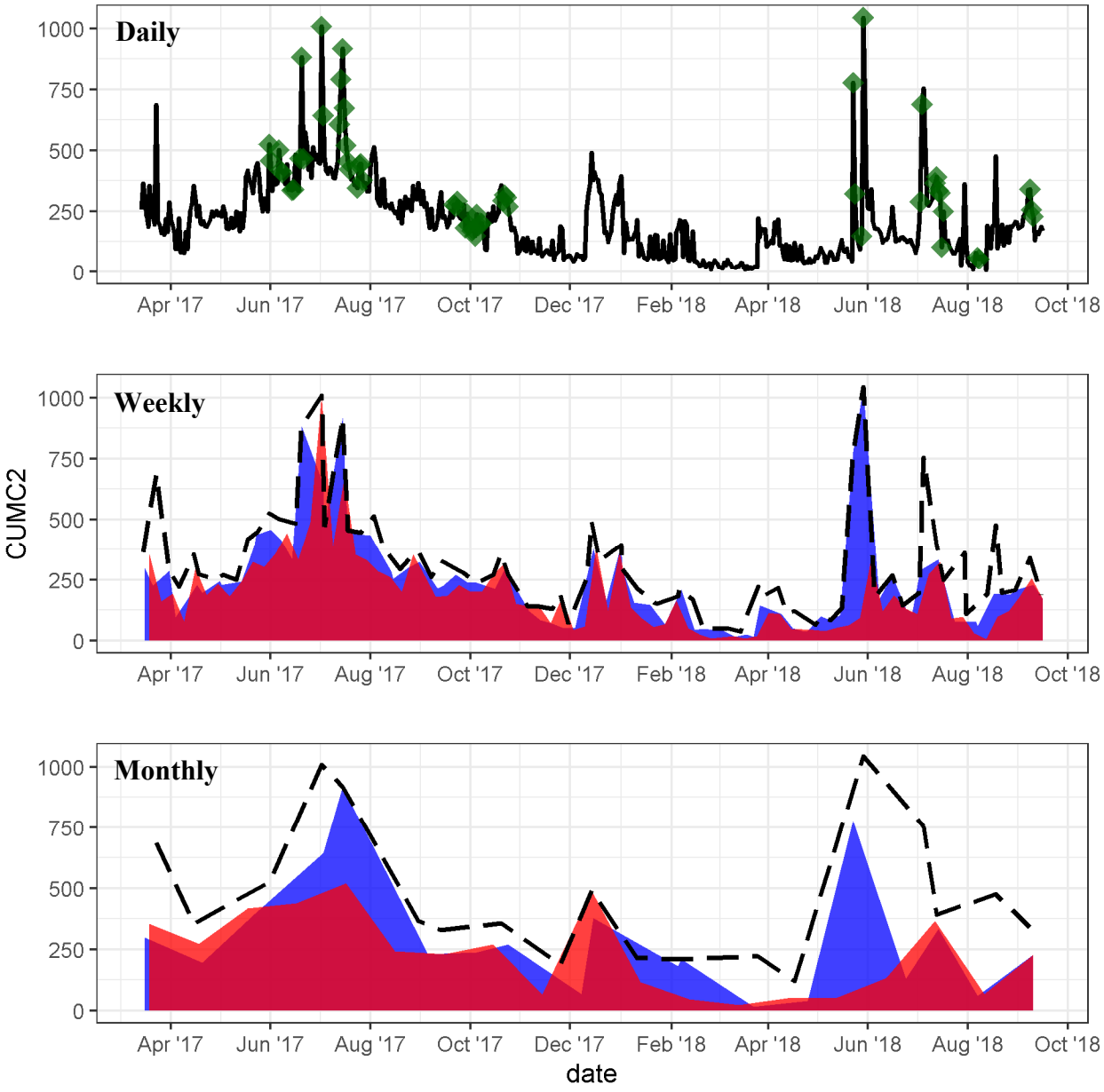



**Figure C64:** Informed vs uninformed sampling strategies for cumulative estimated concentration (CUMC2) in ng·L$^{-1}$; solid black line = average of the imputed datasets; green points = triggered sampling events (≥2); dashed black line = actual representation of risk; blue area = informed sampling, red area = uninformed sampling.

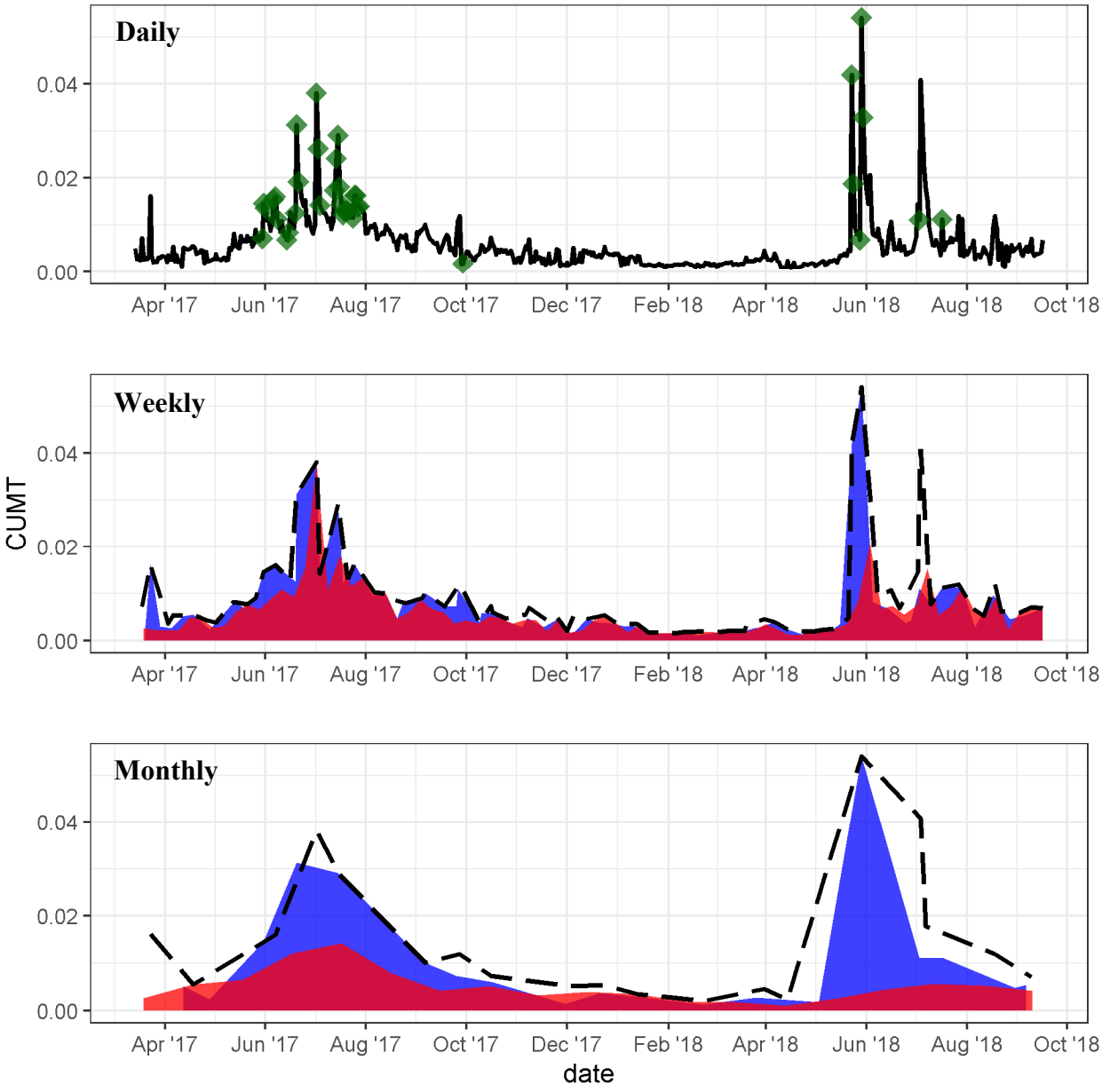



**Figure C65:** Informed vs uninformed sampling strategies for cumulative estimated toxicity (CUMT) expressed as exposure-activity ratio (EAR); solid black line = average of the imputed datasets; green points = triggered sampling events (≥2); dashed black line = actual representation of risk; blue area = informed sampling, red area = uninformed sampling.

# APPENDICES REFERENCES


(1) U.S. Geological Survey. National Water Information System data www.waterdata.usgs.gov/nwis.

(2) Pochodylo, A. L.; Helbling, D. E. Emerging Investigators Series: Prioritization of Suspect Hits in a Sensitive Suspect Screening Workflow for Comprehensive Micropollutant Characterization in Environmental Samples. *Environ. Sci. Water Res. Technol.* **2017**, *3* (1), 54–65. https://doi.org/10.1039/c6ew00248j.

(3) Helbling, D. E.; Hollender, J.; Kohler, H. P. E.; Singer, H.; Fenner, K. High-Throughput Identification of Microbial Transformation Products of Organic Micropollutants. *Environ. Sci. Technol.* **2010**, *44* (17), 6621–6627. https://doi.org/10.1021/es100970m.

(4) U.S. Census Bureau. Area Hydrography. *United States Census Bur.* **2010**.

(5) U.S. Census Bureau. Population and Housing, Population and Housing Unit Counts. *United States Census Bur.* **2012**, No. July, 1–141.

(6) U.S. Census Bureau. National Nation-Level Geography Geodatabase. *United States Census Bur.* **2010**.

(7) U.S. Geological Survey. NED 1/3 Arc-Second. *United States Geol. Surv.* **2015**.

(8) Homer, C. G.; Dewitz, J. A.; Yang, L.; Jin, S.; Danielson, P.; Xian, G.; Coulston, J.; Herold, N. D.; Wickham, J. D.; Megown, K. Completion of the 2011 National Land Cover Database for the Conterminous United States-Representing a Decade of Land Cover Change Information. *Photogramm. Eng. Remote Sensing* **2015**, *81* (5), 345–354. https://doi.org/10.14358/PERS.81.5.345.

(9) U.S. EPA. EPA Facility Registry Service (FRS): Wastewater Treatment Plants www.catalog.data.gov/dataset/epa-facility-registry-service-frs-wastewater-treatment-plants.

(10) Ruttkies, C.; Schymanski, E. L.; Wolf, S.; Hollender, J.; Neumann, S. MetFrag Relaunched: Incorporating Strategies beyond in Silico Fragmentation. *J. Cheminform.* **2016**, *8* (1), 1–16. https://doi.org/10.1186/s13321-016-0115-9.

(11) Menne, M. J.; Durre, I.; Korzeniewski, B.; McNeal, S.; Thomas, K.; Yin, X.; Anthony, S.; Ray, R.; Vose, R. S.; E.Gleason, B.; Houston, T. G. Global Historical Climatology Network - Daily (GHCN-Daily), Version 3. *NOAA Natl. Clim. Data Cent.* **2012**. https://doi.org/10.7289/V5D21VHZ.

(12) Chamberlain, S. rnoaa: "NOAA" Weather Data from R https://cran.r-project.org/package=rnoaa.

(13) NOAA - Northeast Regional Climate Center at Cornell University. Climate Data. 2018.

(14) Pebesma, E. J. Multivariable Geostatistics in S: The Gstat Package. *Comput. Geosci.* **2004**, *30* (7), 683–691. https://doi.org/10.1016/j.cageo.2004.03.012.

(15) Baddeley, A.; Rubak, E.; Turner, R. *Spatial Point Patterns: Methodology and Applications with R*; Chapman and Hall/CRC Press: London, 2015.

(16) Tallaksen, L. M.; Van Lanen, H. A. J. *Hydrological Drought: Processes and Estimation Methods for Streamflow and Groundwater*; 2004; Vol. 48.

(17) Schymanski, E. L.; Jeon, J.; Gulde, R.; Fenner, K.; Ruff, M.; Singer, H. P.; Hollender, J. Identifying Small Molecules via High Resolution Mass Spectrometry: Communicating Confidence. *Environ. Sci. Technol.* **2014**, *48* (4), 2097–2098. https://doi.org/10.1021/es5002105 |.