Environmental Science & Technology

pubs.acs.org/est

Article

Nontarget Chemical Composition of Surface Waters May Reflect Ecosystem Processes More than Discrete Source Contributions

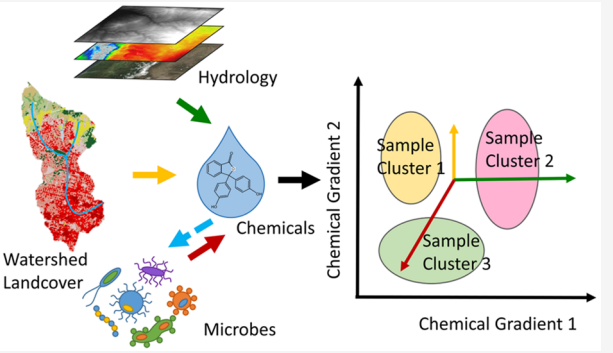
Cheng Shi, Gouri Mahadwar, Emmanuel Dávila-Santiago, Ted Bambakidis, Byron C. Crump, and Gerrad D. Jones*

Cite This: https://doi.org/10.1021/acs.est.2c08540

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: We investigated environmental, landscape, and microbial factors that could structure the spatiotemporal variability in the nontarget chemical composition of four riverine systems in the Oregon Coast Range, USA. We hypothesized that the nontarget chemical composition in river water would be structured by broad-scale landscape gradients in each watershed. Instead, only a weak relationship existed between the nontarget chemical composition and land cover gradients. Overall, the effects of microbial communities and environmental variables on chemical composition were nearly twice as large as those of the landscape, and much of the influence of environmental variables on the chemical composition was mediated through the microbial community (i.e., environment affects microbes, which affect chemicals). Therefore, we found little evidence to support our hypothesis that chemical spatiotemporal variability was related to broad-scale landscape gradients. Instead, we found qualitative and quantitative evidence to suggest that chemical spatiotemporal variability of these rivers is controlled by changes in microbial and seasonal hydrologic processes. While the contributions of discrete chemical sources are undeniable, water chemistry is undoubtedly impacted by broad-scale continuous sources. Our results suggest that diagnostic chemical signatures can be developed to monitor ecosystem processes, which are otherwise challenging or impossible to study with existing off-the-shelf sensors.



KEYWORDS: *nontarget chemical analysis, freshwater microbial communities, spatiotemporal trend analysis, environmental gradient, source attribution, multivariate analysis, watershed analysis, chemical forensics*

INTRODUCTION

Chemical allocation and source identification are central topics in chemical forensic analyses. For decades (e.g., refs 1 and 2), patterns of individual chemical and isotopic tracers have been essential tools for identifying sources of contamination to the environment. Specific examples include pharmaceuticals to detect sewage contamination,[3−6] metals to detect industrial pollution,[7−9] and trace elements to detect hydraulic fracturing fluids.[10] With advances in forensics techniques, tens to hundreds of chemicals features have been used to generate "chemical fingerprints" of different sources (e.g., automotive fluids, stormwater, PFAS/AFFF impacted sites[11−14]). These examples focus on spatially discrete chemical sources of anthropogenic origin. In contrast, spatially continuous sources, which exist across a continuum with no discrete boundaries, are underrepresented within the chemical forensics literature despite contributing to the broad-scale chemical composition of receiving waters.[15−17] This imbalance is likely due to fundamental challenges in collecting and analyzing data generated from continuous chemical sources.

The success of forensics analyses is affected by the spatiotemporal variability of a chemical source. Spatially discrete sources (e.g., stormwater, animal manure) are binary in nature (i.e., present or absent) and readily identifiable at a point in time and space. Therefore, collecting a sample from a discrete source is comparatively easy as long as it is accessible. In contrast, spatially continuous sources exist along a continuum, and there is no obvious point in space that is most representative of these sources, thus making sampling challenging. A strategy for fingerprinting continuous sources could include sampling across a large enough gradient to describe the source of interest. In addition to spatial variability, the chemical composition of continuous sources is likely to be dynamic due to temporal variability in hydrobiogeochemical cycles. Factors such as carbon exchange, solar radiation, soil respiration, streamflow, and others vary hourly, seasonally, and

Special Issue: Data Science for Advancing Environmental Science, Engineering, and Technology

Received: November 14, 2022
Revised: May 8, 2023
Accepted: May 11, 2023

annually;[18−21] thus, the composition of chemicals generated across different gradients are likely to be temporally heterogeneous, even at the same location.

In addition to spatiotemporal variability, chemical fingerprinting of continuous sources is more challenging computationally. In previous work, the success of multivariate chemical fingerprinting decreased as the spatial continuity of the source increased.[22] Multivariate tools used in forensics, such as co-occurrence analysis and classification,[14] gravitate to chemical features that are (largely) unique to a particular source, and unique chemicals were more common for discrete sources.[22] For anthropogenic sources, unique chemicals could arise from manufacturing processes, but even for natural sources (e.g., manure), Dávila-Santiago et al. found a greater proportion of unique features in discrete sources compared to distributed ones (e.g, agricultural runoff and watershed samples).[22] In some instances, discrete sources may consist of greater proportions of labile carbon that contain parent features, whereas continuous sources are likely to contain a greater proportion of byproducts that can be generated from multiple metabolic pathways.[23,24] While fewer unique molecules, particularly small molecules, may exist for continuous sources, we argue that all sources generate distinct chemical gradients that are reflective of the sum of the biogeochemical processes occurring within the source zone. These chemical gradients could be particularly useful for fingerprinting continuous sources when unique chemical features do not exist. We are unaware of forensics studies that probabilistically predict the presence or absence of a source based on the relative position of a sample along a chemical gradient, but such a technique is theoretically possible.

While forensics studies have been able to fingerprint sources and identify their presence within individual samples, methods for identifying the landscape sources from the overall chemical composition of a receiving body of water are generally underdeveloped. With the development of high-resolution mass spectrometry, thousands of nontarget chemical features can be detected in surface water samples, which allows for a semiquantitative assessment of the bulk chemical composition of a water body. Given the diversity of organic molecules, ecosystem processes, and biogeochemical pathways across landscapes, we predict that the probability of two different sources having identical nontarget chemical composition is virtually zero regardless of spatial continuity. Therefore, it should be possible to develop a diagnostic chemical signature associated with any particular source, whether it is spatiotemporally discrete or continuous; however, current forensics strategies are poorly equipped to deal with the spatiotemporal variability in nontarget chemical composition of sources.

We herein report on spatial and seasonal changes in the nontarget chemical composition in riverine systems in the central Willamette Valley, OR. Our goal is to advance computational capabilities in forensics studies and better understand the spatiotemporal drivers of the nontarget chemical composition of surface waters. We hypothesize that the nontarget chemical composition in receiving bodies of water is driven by the most prevalent landscape (spatial) and environmental (temporal) gradients present across a watershed, similar to other studies on sources of dissolved organic carbon (DOC).[25−27] Conversely, autochthonous carbon from autotrophic microbes could structure chemical composition in surface waters.[28,29] To test our hypothesis, we collected water samples from four riverine systems over three seasons. For each watershed, hydrological, weather, and land cover data were collected as environmental variables, combined with microbial community data to explain variation in nontarget chemical composition. Using multivariate computational tools, our specific aim was to identify the drivers structuring the most prominent chemical gradients across all systems. By identifying the specific drivers of these gradients, water quality managers could identify strategies, when necessary, that have the greatest improvement on the holistic chemical composition in receiving bodies of water.

METHODS

Site Descriptions and Watershed Data Collection. Samples were collected on the eastern slope of the Oregon Coast Range in the central Willamette Valley, OR, which has a Mediterranean climate with warm dry summers and cool wet winters with 100−200 cm y^{-1} precipitation.[30] We collected water samples from four rivers (Figure S1). Dixon Creek is ∼10 km long and located entirely within the city of Corvallis. Oak Creek is ∼24 km long and flows from forested headwaters through pastures, grass seed fields, and suburban areas in Corvallis. Marys River is the longest river (∼64 km) in this study and traverses forested headwaters, grass seed fields, and suburban/urban areas in Philomath and Corvallis. Rickreall Creek is ∼40 km long and located about 40 km north of Corvallis. Rickreall Creek traverses forested headwaters, grass seed fields, and urban/suburban area in Dallas, OR.

We collected five longitudinal grab samples (20 L) from each system during summer (June), fall (October), and winter (January, February) 2018/19 (Table S1 for latitude and longitude for each location). Sampling locations were selected at major transitions in land cover type (Figure S1). Each location was approximately equidistant along each creek. By the end of the dry season in fall, Dixon Creek was dry and samples were not collected. We obtained explanatory variables from online databases and field measurements. Upstream watershed characteristics were collected at each point, with land cover percentages acquired from the National Land Cover Database (NLCD, Figure S1, Table S2),[31] and drainage area (Env_Area, square miles), basin slope (Env_Slope, degrees), and mean annual precipitation (Env_AnuPrec, inches) were collected from the USGS StreamStats database (Table S2, methods from[32,33]). For the 2 weeks preceding the sampling event, the hourly precipitation (Env_Rain), baseflow-groundwater runoff (Env_BgRun), and leaf area index (Env_LAI) were averaged to reflect the conditions leading up to the sampling event. Watershed hydrological data were retrieved from the North American Land Data Assimilation System Noah Land Surface Model (NLDAS-NOAH 0.125 × 0.125 degree V2.0[34]). Water pH (Env_pH, unitless) and water temperature (Env_Temp, °C) were collected in the field. Additional details on data collection and preprocessing are in the Supporting Information.

Microbial and Chemical Sampling and Processing. Water samples were split for microbial 16S rRNA gene sequencing and nontarget chemical analysis. Detailed descriptions of DNA sequencing and high-resolution mass spectrometry (HRMS) analyses are in the Supporting Information. Briefly, DNA was extracted and PCR amplified using primers targeting the V4 region of 16S rRNA genes with forward primer 515F and reverse primer 806R[35] according to Earth Microbiome Project Protocols, except that PCR primers were

https://doi.org/10.1021/acs.est.2c08540
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

dual-barcoded following the work of Kozich et al.[36] Genetic data were analyzed using the Quantitative Insights into Microbial Ecology 2 (QIIME2) software.[37] Using QIIME2, the resulting sequences were demultiplexed, filtered, and clustered into operational taxonomic units (OTUs) at 97% sequence similarity. Final taxonomic identity was determined by comparing OTUs against the Greengenes database.[38] HRMS analysis and data processing were described previously[22] and are summarized in the Supporting Information. Briefly, triplicate samples (1 L each) were filtered, spiked with deuterated internal standards, and loaded onto C18 SPE cartridges (1,000 mg of sorbent, Resprep). Samples were eluted with methanol, dried, and resuspended in 90:10 (v/v) water:methanol for LC-HR-TOF-MS (Shimadzu Nexera UHPLC coupled to an AB Sciex 5600 tripleTOF). HRMS analysis as well as peak picking and retention time alignment (Sciex Peakview v2.2 and Markerview v1.3.1) were performed at Oregon State University's Mass Spectrometry Center. Instrument performance and quality control procedures are introduced in the Supporting Information. Overall, 12,638 microbial OTUs were identified in 46 samples and 1,605 chemical features were retained from 55 samples (Table S1 for microbial data availability).

Hypothesis Testing and Data Analysis. We hypothesized that the nontarget chemical composition in rivers is driven by the most prevalent landscape and environmental gradients present across watersheds.[25−27] Accordingly, we made the following predictions. First, sample locations with similar surroundings and/or upstream land cover will be more chemically similar and will cluster together in all seasons. Second, landscape influences on the chemical composition will be more pronounced during the winter rainy season when overland runoff increases the chemical loads to surface waters. Finally, we expected landscape and environmental variables to explain the greatest amount of variability within the chemical data set. If these patterns are observed, it suggests that the dominant chemical gradients present in riverine systems can be attributed to processes occurring across the landscape.

We evaluated these predictions using multiple tools. Qualitatively, we used nonmetric multidimensional scaling (NMS) to evaluate spatiotemporal clustering of the chemical data (PC-ORD v. 7.0.2[39]), which evaluates our first and second predictions. Instead of analyzing all 1,605 chemical features and all 12,638 OTUs individually, NMS was used to reduce the dimensionality of the data to a smaller number of composite dimensions that summarize the information contained within the original data set. NMS is not constrained by assumptions of multivariate normality, linearity, and homoscedasticity,[40] which is advantageous when working with high-dimensional data. The raw chemical data were logarithmically transformed ($\acute{x} = \log_{10}(x + 1)$) before import into NMS analysis to reduce the influence of outliers and help achieve normally distributed data for other analyses. NMS analyses were performed using the Bray−Curtis distance, which is appropriate and commonly used for sparse, high-dimensional, and bounded matrices.[39] A randomization test was used to determine the appropriate number of NMS dimensions to retain. The correlation between NMS axes and explanatory variables was used to help interpret each NMS axis and identify potential explanatory variables. Complete details on the NMS analysis and its assumptions are presented in the Supporting Information.

We were able to amplify microbial DNA from 46 out of the 55 samples with chemical data; therefore, analysis with the microbial data consisted of a subset of the entire chemical data set. To illustrate the microbial community distribution, we used hierarchical cluster analysis with Bray−Curtis distance and average linkage method, using the "clustermap" function of the seaborn package (v.0.11.1) in Python to display the relationships among microbial orders. The assumptions of this method are the same as NMS (Supporting Information). To explore the source of microbial populations, we used the text mining software SEQenv[41] (v.1.3.0) to identify sequence matches (97% identity) for all OTUs in the NCBI nucleotide (nt) database. Matched sequences were used to extract textual metadata and parse terms associated with Environmental Ontology controlled vocabulary in the literature. Default settings were used except we set "−min_identity 0.97" and "−max_targets 50". We manually classified terms into several categories (Table S3) and added sequences not classified by SEQenv to the output table categorized as "unclassified."

We used canonical correspondence analysis (CCA) to quantify the variability in both chemical and microbial data sets explained by their associated predictor variables (PC-ORD v. 7.0.2;[39]), which evaluates our third prediction. Unlike NMS, CCA is a direct gradient analysis method and only considers the structure of a matrix that is related to predictor variables of interest.[42] We grouped all predictor variables into three broad categories including environmental (e.g., water temperature, pH; $n = 8$), land cover ($n = 7$), microbial NMS axes ($n = 2$; used when predicting chemicals), and chemical NMS axes ($n = 3$, used when predicting microbes). Because of the unequal number of variables within each subset, we calculated the adjusted R-square values based on a permutation procedure to reduce statistical bias and to better compare the relative contribution of each variable type.[43] More details on CCA assumptions and adjusted methods are in the Supporting Information.

Finally, we used structural equation modeling (SEM) to quantify the relationship between the predictor variables and chemical data, which also addresses our third prediction. SEM is a causal modeling approach that evaluates the statistical significance of a hypothesized network, which is developed *a priori*, and its drivers. Unlike ordination, SEM allows us to separate direct, indirect, and total effects of predictor variables on dependent variables. Additionally, SEM also allows us to evaluate feedback loops, which cannot be addressed with CCA. Specifically, microbes could structure the chemical composition, which could in turn structure the microbial community. SEM analysis was performed using SPSS AMOS (v25) with chemical and microbial NMS axes and summarized environmental and land cover variables as predictors (see the Supporting Information for complete details on SEM).

RESULTS AND DISCUSSION

Spatiotemporal Variation in Nontarget Chemical Composition. Following NMS on the chemical data, a stress of 13.79 was achieved using 3 axes (Figures 1, S2, and S3). The minimum stress for randomized data was 26.54 (Figure S2), indicating that the strong structure within the original data was not due to random chance ($p = 0.02$). A 3-dimensional ordination was deemed sufficient because the additional reduction in stress between axes 3 (i.e., 13.79) and 4 (i.e., 9.93) was small (Figure S2) and a final stress <20 is considered a good fit (Supporting Information).[39] The final configuration

https://doi.org/10.1021/acs.est.2c08540
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

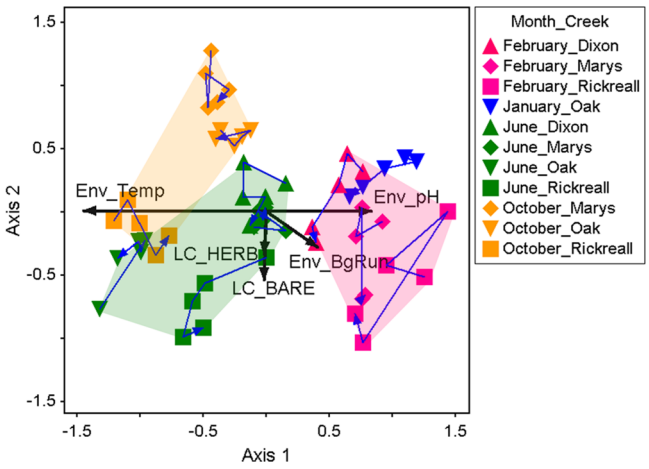


Figure 1. Nonmetric multidimensional scaling (NMS) ordination of nontarget chemical composition (Axes 1 and 2) in water samples, color coded by sampling months, and symbol coded by watersheds (Creek). The blue successional vectors indicate sample locations from upstream to downstream. The overlaid black arrows indicate environmental variables with an *r*-square cutoff value of 0.15.

accounted for 63.1% of the total variation in the original data set by the Bray−Curtis distance, where axes 1, 2, and 3 summarize 32.5%, 15.9%, and 14.6% of variation, respectively. Chemical NMS-1 was most correlated with seasonal explanatory variables including water temperature (Env_Temp), pH (Env_pH), and baseflow-groundwater runoff (Env_ BgRun) ($r_{pearson} > 0.4$, Table S2, Figure 1). Chemical NMS-2 is correlated with barren (LC_BARE) and herbaceous (LC_HERB) land cover ($r_{pearson} \geq 0.39$, Table S2, Figure 1). Chemical NMS-3 is correlated with pH (Env_pH) and vegetation coverage (Env_LAI) ($r_{pearson} > 0.4$ in Table S2; Figure S3).

We predicted that the chemical composition would cluster by land cover, but we found no evidence to support this prediction. Instead, samples clustered mainly by season in the NMS ordination. Winter samples (positive values) separated from summer and fall samples (mostly negative values) along chemical NMS-1, while fall samples (mostly positive values) separated from summer and winter samples (mostly negativevalues) along chemical NMS-2 (Figure 1). Chemical NMS-1 appears to be a seasonal chemical gradient that is strongly related to temperature. The nontarget chemical composition of samples collected in summer and fall overlaps almost entirely on chemical NMS-1. If chemical NMS-1 is structured by temperature, this overlap may be expected given that there was no statistical difference in the average daily temperature in the week leading up to the sampling events (June: $\mu = 15.4$ °C, $\sigma = 1.8$; October: $\mu = 14.4$ °C, $\sigma = 2.8$, two sample *t* test assuming unequal variance: $t = 1.61$, df = 43, $p = 0.11$). We found evidence of temporal clustering on chemical NMS-1 but no evidence of clustering by land cover type; however, we can provide no mechanistic explanation for temporal clustering by examining the NMS plot.

We predicted that changes in nontarget chemical composition would shift as each stream traversed major land cover types; however, we found limited evidence to support this prediction. Instead, chemical variability between samples within each river increased within each river system increased as NMS-2 became more negative (Figure 1). That is, the longitudinal (along-channel) change in the chemical composition between samples within each river was small in the fall (positive values on chemical NMS-2) but large in the summer and winter (negative values on chemical NMS-2). These results are consistent with allochthonous inputs from overland runoff in general but not necessarily from specific landcover types.[44] During the winter rainy season when the vadose zone is virtually saturated within the Willamette Valley, and even at the beginning of the dry season (i.e., June), chemicals present in overland runoff and soil seepage readily discharge into surface waters. We propose that this chemically enriched water drove negative values on chemical NMS-2. Conversely, the most positive values on chemical NMS-2 are from October Marys River and Oak Creek. In October, Dixon Creek was dry, so no comparison could be made. October is at the end of the dry season within the Willamette Valley, so flow is maintained by groundwater. Furthermore, the chemical data set consists of nonpolar organics, which are expected to absorb to soil and removed from the water column.[45,46] Finally, phototrophic organisms could generate autochthonous inputs that maintain a relatively constant chemical composition during low flow periods. Although other tools are needed (e.g., hydrophobic organic acid fraction[47]) to distinguish allochthonous and autochthonous carbon, a likely explanation of the variability along chemical NMS-2 is that seasonal changes in hydrology drive chemical variability during wet periods but help stabilize the chemical composition during dry periods.

We have preliminary evidence to support our *ad hoc* hypothesis that chemical enrichment is driving chemical NMS-2. If true, we predict inverse correlations between chemical NMS-2 and chemical richness (i.e., the number of chemical features present), Shannon diversity, and mean feature intensity, which was observed ($n = 55$; $r_{spearman} = -0.61$, $p < 0.01$; $r_{spearman} = -0.61$, $p < 0.01$; $r_{spearman} = -0.57$, $p < 0.01$; respectively; Figure S4). In addition, the influence of wastewater supports our chemical enrichment assessment of chemical NMS-2. Unlike October Marys River and Oak Creek, which were longitudinally chemically homogeneous in October, Rickreall Creek exhibits some longitudinal chemical change (Figure 1), particularly at locations 4 and 5 in all seasons. These locations are influenced by wastewater from the city of Dallas, which discharges treated effluent between locations 3 and 4. In all three seasonal clusters, Rickreall Creek sampling locations 4 and 5 are the most negative samples on chemical NMS-2. Interestingly, treated wastewater effluent is also discharged into the Marys River between locations 3 and 4, but this only occurs during high flows the winter due to discharge permit requirements. Accordingly, Marys River sampling locations 4 and 5 during the winter exhibit strong negative values on chemical NMS-2. These chemical additions appear to drive negative trajectories on chemical NMS-2, which supports our *ad hoc* hypothesis. Overall, we find little evidence to suggest that landcover variation contributes to the chemical composition of rivers on NMS-2. Because we focus on small streams to capture locally generated chemicals, it is possible the landcover gradient is not long enough to exert a detectable change in the chemical composition.

Finally, we found evidence to suggest that a small proportion of the chemical composition was unique to each system based on clustering along chemical NMS-3 (Figure S3). Marys River samples clustered on negative values of NMS-3, Oak Creek samples clustered around 0, and Rickreall Creek clustered on positive values. Dixon Creek exhibited the most seasonal variability on NMS-3, which was most strongly correlated with

https://doi.org/10.1021/acs.est.2c08540
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

LAI. Dixon Creek differs from the other rivers because it is entirely (sub) urban and its watershed is dominated by deciduous ornamental trees. Therefore, seasonal variability in LAI is expected to be extreme due to leaf drop in fall and leaf growth in spring. The chemical variability represented by chemical NMS-3 may be related to seasonal decomposition of leaves,[48] variability in root exudates,[49−51] or dynamic leaching.[52] Although the mechanism is unknown, deciduous trees within a watershed may explain some variability in the chemical composition of these streams.

Previous studies indicated that anthropogenic chemicals are significant contributors to water quality degradation in rivers because river waters transport contaminants from agricultural and urban areas through watersheds.[53−57] While specific pollutants present in rivers are certainly derived from the landscape,[54,58] targeted chemical analyses focused on anthropogenic contaminants eliminate signals associated with environmental processes and ignore their contribution to water chemistry and ecosystem health. Our results suggest that the dominant chemical gradients present in receiving waters are driven by seasonal temperature, hydrological, and perhaps phenological processes. Seasonal variation of nontargeted chemical composition in small rivers was also reported in previous research.[57,59] While the NMS analysis provides little mechanistic information, it highlights that the bulk chemical composition in waters is dynamic. As a result, periodic (e.g., quarterly) water quality sampling may not be sufficient to identify the mechanisms driving particular water quality phenomena.

Testing Causal Drivers of the Chemical Structure. Canonical correspondence analysis (CCA) was used to determine the variation in the chemical data set that can be explained by each variable. We used the microbial NMS axes as explanatory variables for the chemical variance (more details below in Microbial Community Interpretation). After controlling for differences in the number of variables, the environmental set explained 8.3% of variation in the chemical data set (inertia) by 3 canonical axes, followed by the microbial set (5.2% by 2 axes) and land cover set (3.9% by 3 axes). As with the NMS assessment, the CCA results suggest that land cover plays a smaller role in structuring the nontarget chemical composition than the environmental and microbial variables. While microbial communities may contribute to the structure of the nontarget chemical composition, it is possible and even likely that environmental and land cover variables structure the microbial community and thus indirectly drive the nontarget chemical composition. The adjusted CCA results suggest that environmental and land cover variables explain 8.0% and 3.4% of the variation in the microbial communities, respectively. These results indicate that environmental variables are drivers of the microbial community, which is typical of freshwater communities.[60] In addition, 6.7% of adjusted variation in microbial communities can be explained by the chemical NMS axes. Although these results suggest that a feedback loop might exist where microbial communities affect the chemical composition which in turn drives the microbial communities, CCA is not capable of characterizing this type of interaction. Feedback loops were further explored using SEM. Nevertheless, the large fraction of unexplained variation in these analyses reveals the complex nature of nontarget chemical composition in river waters.

The final SEM model is depicted in Figure 2. All connections retained within the SEM analysis were statistically

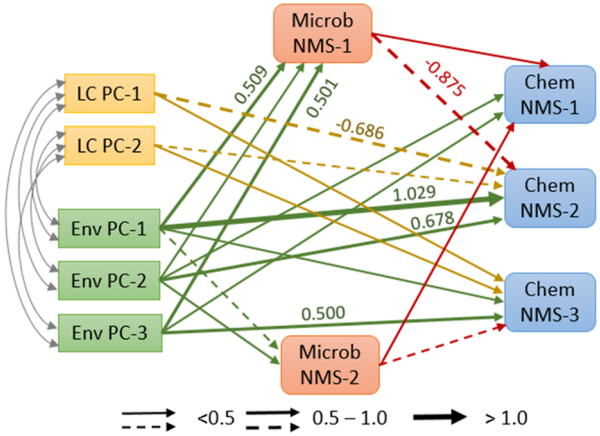


Figure 2. Final structural equation model with estimated weights on important pathways. Line width is scaled by the standardized regression coefficients, which were labeled if they are greater than 0.5. Solid lines represent positive direct effects, and dashed lines represent negative direct effects. LC PCs 1 and 2 are two principal components representing 7 land cover variables. Env PCs1−3 are principal components representing 8 environmental variables. Microbial NMS-1 and 2 are scores on NMS ordination axes of microbial communities. Chemical NMS-1−3 are scores of nontarget chemical NMS ordination.

significant (i.e., $p < 0.05$). The Chi-square ($\chi^2 = 28.5$, df = 20, $p = 0.1$) and most fit indices were at or above critical thresholds (CFI = 0.96, NFI = 0.90, GFI = 0.91, TLI = 0.92, Table S4) indicating that the hypothesized model is a good fit and statistically meaningful.[61,62] While the RMSEA was above our desired threshold (0.097 > 0.08), this fit was near the threshold, and with the other fit indices indicating a good model fit, we were confident that the SEM model results were robust. Overall, the model best characterized the variability of chemical NMS-1 ($R^2 = 0.79$), microbial NMS-1 ($R^2 = 0.76$), and chemical NMS-2 ($R^2 = 0.60$). The model performance was moderate for microbial NMS-2 ($R^2 = 0.44$) and chemical NMS-3 ($R^2 = 0.41$).

The standardized total, direct, and indirect effects estimated in the SEM model are presented in Table S5, and only standardized effects are discussed herein. Chemical NMS-1 was most strongly driven by Env PC-2 (i.e., baseflow-groundwater runoff) and Env PC-3 (i.e., temperature) and weakly by microbial NMS-1 and -2. Env PCs-1, -2, and -3 had a moderate direct effect on microbial NMS-1, which added indirect effects on chemical NMS-1. Overall, the positive direct effects from Env PC-2 (i.e., high baseflow and groundwater runoff) and Env PC-3 (i.e., low temperature) to chemical NMS-1 are consistent with the chemical NMS analysis discussed previously. Finally, microbial NMS-1 and 2 (see Microbial Community Interpretation for details on the microbial community) also have positive direct effects on chemical NMS-1. Chemical NMS-2 was strongly driven by Env PCs-1 (i.e., LAI) and -2 (i.e., baseflow-groundwater runoff), microbial NMS-1, and LC PC-1 (urban to forest gradient). Unlike with chemical NMS-1, environmental variables exerted moderate to strong effects on chemical NMS-2 and microbial NMS-1, but the microbial NMS-1 has strong negative direct effects on chemical NMS-2. Therefore, the total effects of Env PCs-1 and -2 were moderated by their indirect effect mediated through microbial NMS-1. Overall, Env PCs-1 and -2 have positive direct effects, and LC PCs-1 (i.e., low forests) and -2

https://doi.org/10.1021/acs.est.2c08540
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

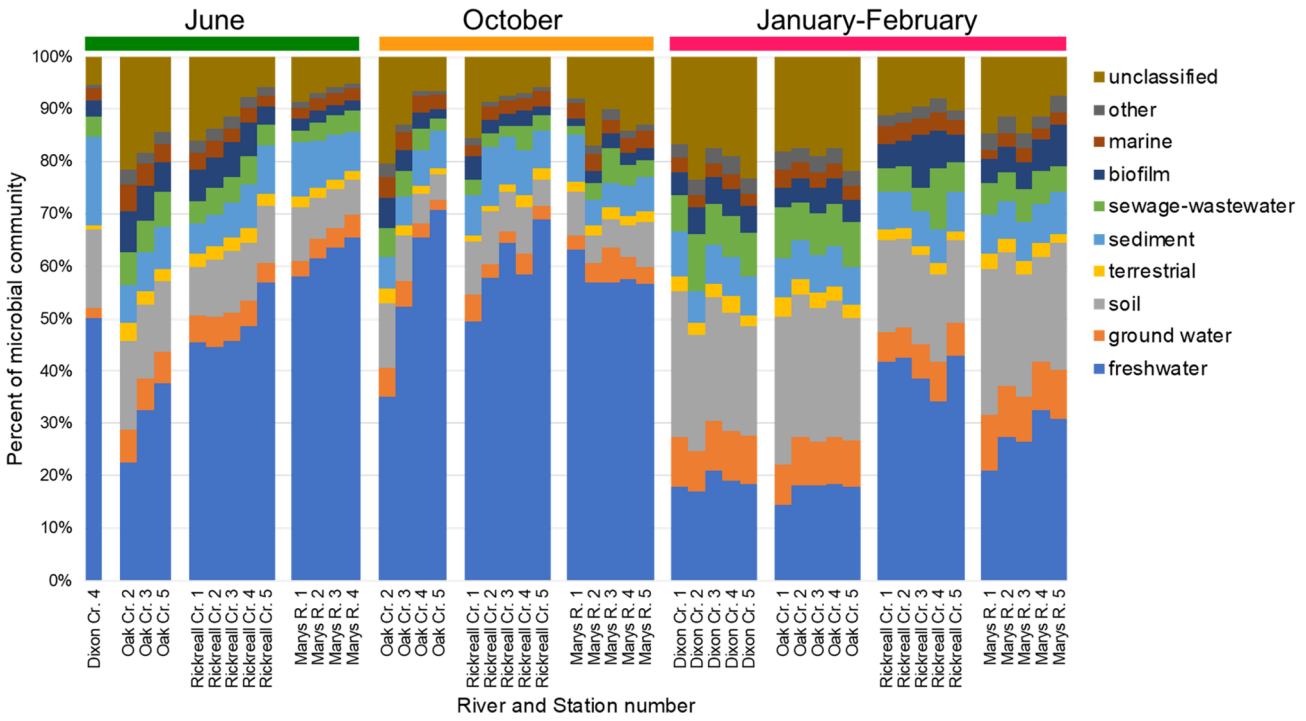


Figure 3. Relative contribution of predicted environmental sources to microbial communities based on SEQenv text mining of sources for related taxa. Samples are grouped by season and by creeks in each season to illustrate the microbial changes from upstream to downstream.

(i.e., high agriculture) have negative direct effects on chemical NMS-2. Chemical NMS-3 was most strongly driven by Env PC-3 (i.e., temperature), LC PC-1 (i.e., developed land to forest gradient), and Env PC-1 (i.e., LAI) which is consistent with the NMS results. All other variable interactions were weak and not discussed herein.

We considered the possibility that the nontarget chemical composition could influence the microbial community composition, resulting in a feedback loop. The SEM results suggest that the microbial communities helped structure the chemical composition, yet we know from the literature that the chemical composition affects the microbial community.[63,64] We tested this hypothesis with multiple nonrecursive SEM models, but all permutations resulted in poor model fit. Therefore, the observed pattern in the CCA analysis (i.e., the chemical composition explains the microbial variability) is likely a result of correlation between the two data sets instead of causation. The SEM result does not necessarily mean that chemicals do not structure the microbial community, but it highlights that the microbial communities within this system may be more sensitive to changes in environmental conditions than to the chemical composition.

Research in DOC chemistry has demonstrated that land use, microbial activities, and seasonality are important drivers to changes in DOC characteristics in river waters.[65−67] We also found their significant contribution to nontarget chemical composition. Kothawala et al. suggested that land use might have greater influence in changes in DOC chemistry than seasonality.[27] However, they only used one categorical variable as seasonality to explain five DOC components, which might be too simplified to quantify the seasonal contribution. Even though our land cover data are continuous (i.e., percentages), the results are derived from static NLCD data. This might explain why land cover data was less important in the CCA and SEM analysis, but there was no discernible land cover pattern in the NMS analysis, suggesting that the absence of a strong land cover signal in the former analyses is not an artifact of the data type.

Our interpretation of the SEM model is based on the extracted gradients, whereas the interpretation of the chemical NMS results is based on metadata that are not readily captured by the land cover and environmental variables. As a result, some of the conclusions drawn from the SEM results, particularly related to chemical NMS-2, neither support nor contradict the interpretation of the NMS results. The SEM results suggest that microbial communities interact strongly with environmental conditions to structure the nontarget chemical composition in streams. This does not imply that landscape and land cover are unimportant, especially considering that chemical NMS-2 is influenced by hydrology. Based on the large direct effects of the microbial axes and their strong intermediary effects, the microbial community is arguably the most important variable structuring the nontarget chemical composition in streams.

Microbial Community Interpretation. Microbial axes were statistical drivers of the nontarget chemical composition; however, our statistical analyses could not determine if the microbial community is a mechanistic driver of the nontarget chemical composition (e.g., via metabolic processes) or if their co-occurrence comes from a shared but uncharacterized environmental driver. Therefore, we explored patterns within the microbial community for additional insights into structural drivers. NMS ordination of the microbial data yielded a stress of 12.20 using 2 axes (Figure S5). The minimum stress for randomized data was 21.93 (Figure S2), indicating that the strong structure within the original data was not due to random chance (p = 0.02). The final configuration represents 86.3% of total variation of the original data, with 78.9% and 7.4% for NMS axes 1 and 2, respectively. Therefore, a 2-

https://doi.org/10.1021/acs.est.2c08540
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

dimension configuration was deemed adequate to represent the microbial community (Figure S5).

In general, winter communities and summer/fall communities were separated along microbial NMS-1; however, there was no clear separation in 2 dimensions for the summer and fall samples (Figure S5). Overall, microbial NMS-1 was most correlated with water temperature (Env_Temp), pH (Env_pH), and vegetation cover (Env_LAI). Nearly all creeks exhibited the same spatiotemporal trend: downstream samples became increasingly negative on microbial NMS-1. This suggests a consistent pattern of microbial community change from upstream to downstream along each creek, potentially driven by changes in hydrology, vegetation cover, and river chemistry. For the same creek, trajectories of microbial community change were similar for summer and fall samples but differ in winter samples (Figure S5). Microbial communities also varied weakly by river along microbial NMS-2, with Marys River samples on the positive side of the axis, and Oak and Dixon creeks on the negative side. Microbial NMS-2 is also influenced by hydrology (Env_Rain, Env_BgRun), which appear to cause the winter samples to be more positive on microbial NMS-2. Since microbial NMS-1 represents the largest microbial variation and has greater impacts on chemical variation from SEM results, we focused on microbial NMS-1 for further analysis.

The strongest signal of variability in microbial communities was seasonal differences between winter communities and summer and fall communities along microbial NMS-1, matching changes in chemical composition along chemical NMS-1 (R = 0.64, p < 0.01). There were also moderate correlations between microbial NMS-2 and chemical NMS 1 and 3 (Figure S5). Seasonal variability is common for river microbial communities, which form annually repeating seasonal assemblages.[68] To explore seasonal changes in microbes we summed relative abundances of OTUs classified into the 20 most abundant orders (Figure S6) and found these formed three groupings based on occurrence and abundance across all samples. The two groups more strongly associated with winter samples included orders typical of soils and groundwater including *Clostridiales*, *Legionellales*, and *Methylophilales*. The group more strongly associated with summer and fall samples featured orders that include many typical freshwater planktonic bacteria such as *Burkholderiales* and *Actinomycetales*.

We found strong patterns in the predicted sources of OTUs based on SEQenv analysis (Figure 3). The greatest variability occurred for OTUs associated with soil, groundwater, and freshwater sources. Winter samples featured higher proportions of soil and groundwater taxa, suggesting a terrestrial source for many of the microbes in these samples. This indicates rainy conditions in winter lead to greater surface runoff and soil porewater contributions to the microbial communities. A similar pattern was found in winter for chemical diversity (Figure 1). In contrast, summer and fall samples featured much higher proportions of freshwater taxa, suggesting that higher temperatures and lower flow rates promoted the growth of planktonic microbes in these rivers. These planktonic communities may be responsible for structuring the nontarget chemical composition in summer and fall. Moreover, the proportions of freshwater, soil, and groundwater taxa exhibited small changes from headwaters to downstream sites during the winter, indicating a continued influence of terrestrial microbes along the flow path. Unlike winter samples, the proportion of freshwater taxa increased along the flow path of each river in summer and fall, suggesting that the consistent trajectories of change in microbial communities identified in Figure S5 are caused by the development of a planktonic microbial community in each river.

Similar sourcing information is not available for the chemicals identified with mass spectrometry in this study, but the correlation between microbial communities and organic chemicals suggests strong interactions between these two complex assemblages. Links between microbes and organic chemistry in aquatic environments have been discussed in previous studies.[69,70] Microbes in river waters modify and metabolize organic molecules and excrete secondary metabolites into the environment. In microbial ecology, researchers have reported that microbial community structure is shaped by organic carbon inputs into aquatic systems.[71] However, from our SEM results, we did not find statistically significant direct effects from chemical gradients to microbial gradients. In contrast, our results suggest that microbial communities respond to the environmental changes directly, causing them to change their behavior or community structure to adapt to the environmental conditions. Shifts in community subsequently modify the organic chemical composition by transforming molecules and excreting different suites of metabolites into the environment. Microbial communities in winter likely correspond to terrestrial organic carbon inputs from groundwater/soil seepage, while communities in summer are likely driven by environmental changes in aquatic ecosystems as planktonic microbial communities develop. The temporal patterns in microbial communities and nontarget chemicals suggest different underlying mechanisms of microbial interaction with nontarget chemicals in different seasons, which can be useful for evaluating ecosystem health and the effects of environmental/land cover changes at ecosystem scales.

The primary objective of this study was to identify the environmental drivers structuring the nontarget chemical composition in watersheds. We hypothesized that the dominant spatial gradients across a watershed would drive the chemical composition, but temporal trends were more prevalent than spatial ones. Our results indicate that the temporal variability in the nontarget chemical composition was due to seasonal changes in ecosystem processes, specifically shifts in microbial communities, changes in runoff, and perhaps phenology. Microbial communities are the most diverse members of an ecosystem and are involved in almost all processes that produce organic molecules; therefore, it is not surprising that their signal was so prominent. However, this conclusion is likely biased toward limitations in this study, as discussed above. Large portions of variation (48.5%; Figure S7) are unexplained by the explanatory covariates discussed in this paper. The interaction between sediments and river water, the contribution of aquatic organisms, and many other associated contributors to the nontarget chemical composition in watersheds should be investigated in the future. Incorporating analysis of water chemistry including total organic carbon, DOC optical properties, nutrient loading, and macroinvertebrates could be very useful for identifying causal relationship between chemical variation and ecosystem processes. These efforts will ultimately help us better understand the ecosystem processes that structure the nontarget chemical composition and design management frameworks to improve ecosystem health and water quality. Future studies to develop forensic workflows and identify diagnostic chemical features of specific ecosystem processes are needed to track temporally and

https://doi.org/10.1021/acs.est.2c08540
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

spatially distributed chemical sources. Along with ordination and other gradient analyses, data filtering, normalization (e.g., by runoff volumes or flow), transformation (e.g., Fourier transformation) could be used to filter dominant signals that represent specific environmental gradients of interest.

LIMITATIONS

Various data considerations exist that limit our ability to make definitive conclusions about the factors structuring the nontarget chemical composition in these systems. First and foremost, all multivariate tools have limitations, and it is necessary for all researchers to assess whether a data set is appropriate for a particular tool. We checked all assumptions (see Supporting Information for details), and most tools were deemed suitable. The unimodal assumption for CCA was certainly violated for some microbial OTUs and chemical features. While CCA is a robust tool, a prudent compromise is to relax the assumptions to facilitate usage but recognize that this weakens the strength of the results. Therefore, we used CCA as a qualitative tool (i.e., assessing the relative importance of variables to chemical and microbial matrices) instead of a quantitative one (i.e., using the ordination axis in an SEM regression). Second, samples in this study were limited in sampling frequency and geospatial extent. While seasonal sampling was enough to detect variability, more samples are needed to evaluate the breadth and dynamism of changes in the chemical composition and the mechanisms driving that change. Furthermore, all the watersheds are in the same region with similar climatic and annual hydrological conditions. It is possible that the spatial extent was not large enough to capture chemical variability driven by land cover heterogeneity. Third, although land cover could be useful for describing chemical sources, the NLCD land cover data collected from StreamStats is static and thus cannot explain temporal variability within the chemical data set. While dynamic variables (e.g., leaf area index) could be useful for describing temporal variability in the chemical composition, they are not necessarily attributable to specific chemical sources (e.g., deciduous vs pine forests). It is possible that more variability in the chemical composition could be attributed to landscape processes if temporal data could be linked with spatially explicit sources and vice versa. Finally, we are limited by the SPE and HRMS methods used in this study, where only a subset of chemicals in the organic carbon pool was collected and analyzed. We used C18 SPE cartridges in our analysis, which bias our results to nonpolar molecules. Other SPE cartridges could be used instead of—or in tandem with—C18 to obtain a broader suite of nontarget chemical features. In addition, our results are biased by our HRMS data preprocessing methods. We tried to be as conservative as possible when retaining features to avoid false positive (Supporting Information). Our rational is that chemical gradients used for forensics purposes should contain as few false positives as possible, but our conservative approach likely increase the incidence of false negatives. Furthermore, we are uncertain if restricting our m/z range to small molecules (i.e., $m/z < 1{,}000$) predisposes our analysis to detect gradients generated by microbes compared to landscapes contributions, which may have larger molecular weight compounds that are outside our detection range. It is important to recognize and acknowledge that chemical extractions and MS techniques will always bias the results. Further research needs to be done to examine how methodological approaches bias the chemical gradients that can be extracted from the environment.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.est.2c08540.

Detailed information on sampling sites, chemical and microbial sample processing steps, instrument performances, data processing and evaluations, statistical methods and justifications, and additional results in the form of tables and figures (PDF)

AUTHOR INFORMATION

Corresponding Author

Gerrad D. Jones − *Department of Biological & Ecological Engineering, Oregon State University, Corvallis, Oregon 97331-4501, United States;* orcid.org/0000-0002-1529-9506; Email: gerrad.jones@oregonstate.edu

Authors

Cheng Shi − *Department of Biological & Ecological Engineering, Oregon State University, Corvallis, Oregon 97331-4501, United States;* orcid.org/0000-0001-6985-1514

Gouri Mahadwar − *Department of Biological & Ecological Engineering, Oregon State University, Corvallis, Oregon 97331-4501, United States*

Emmanuel Dávila-Santiago − *Department of Biological & Ecological Engineering, Oregon State University, Corvallis, Oregon 97331-4501, United States*

Ted Bambakidis − *Department of Microbiology, Oregon State University, Corvallis, Oregon 97331, United States*

Byron C. Crump − *College of Earth, Ocean, and Atmospheric Sciences, Oregon State University, Corvallis, Oregon 97331, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.est.2c08540

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We acknowledge that Oregon State University in Corvallis, Oregon, is located within the traditional homelands of the Marys River or the Ampinefu Band of Kalapuya. Following the Willamette Valley Treaty of 1855, Kalapuya people were forcibly removed to reservations in Western Oregon. Today, living descendants of these people are a part of the confederated Tribes of Grand Ronde Community of Oregon (grandronde.org) and the confederated tribes of the Siletz Indians (ctsi.nsn.us). We acknowledge Rebecca Hutchinson, Jeffrey Jenkins, Claudia Maier, Jeff Morre, Lewis Semprini for their assistance. We thank 3 anonymous reviewers who provided constructive feedback during the review process. We acknowledge the J. Ron Miner Graduate Fellowship for Water and Environmental Stewardship for its financial support. Funding was provided by the National Science Foundation (NSF, CBET 1949013, EAR 1836768, DEB 1840243) and Oregon State University-Agricultural Research Service (#8986A). The authors declare no financial conflict of interest.

REFERENCES

(1) Hawthorne, S. B.; Krieger, M. S.; Miller, D. J.; Mathiason, M. B. Collection and quantitation of methoxylated phenol tracers for

atmospheric pollution from residential wood stoves. *Environ. Sci. Technol.* **1989**, *23* (4), 470−475.
(2) Lindau, C. W.; Delaune, R. D.; Patrick, W. H.; Lambremont, E. N. Assessment of stable nitrogen isotopes in fingerprinting surface water inorganic nitrogen sources. *Water, Air, and Soil Pollution* **1989**, *48* (3), 489−496.
(3) Van Stempvoort, D. R.; Roy, J. W.; Brown, S. J.; Bickerton, G. Artificial sweeteners as potential tracers in groundwater in urban environments. *Journal of Hydrology* **2011**, *401* (1), 126−133.
(4) 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* **2017**, *9* (2), 143.
(5) McCance, W.; Jones, O. A. H.; Edwards, M.; Surapaneni, A.; Chadalavada, S.; Currell, M. Contaminants of Emerging Concern as novel groundwater tracers for delineating wastewater impacts in urban and peri-urban areas. *Water Res.* **2018**, *146*, 118−133.
(6) Sankararamakrishnan, N.; Guo, Q. Chemical tracers as indicator of human fecal coliforms at storm water outfalls. *Environ. Int.* **2005**, *31* (8), 1133−1140.
(7) Hansmann, W.; Köppel, V. Lead-isotopes as tracers of pollutants in soils. *Chem. Geol.* **2000**, *171* (1), 123−144.
(8) Gao, B.; Zhou, H.; Liang, X.; Tu, X. Cd isotopes as a potential source tracer of metal pollution in river sediments. *Environ. Pollut.* **2013**, *181*, 340−343.
(9) Sivry, Y.; Riotte, J.; Sonke, J. E.; Audry, S.; Schäfer, J.; Viers, J.; Blanc, G.; Freydier, R.; Dupré, B. Zn isotopes as tracers of anthropogenic pollution from Zn-ore smelters The Riou Mort-Lot River system. *Chem. Geol.* **2008**, *255* (3), 295−304.
(10) Warner, N. R.; Darrah, T. H.; Jackson, R. B.; Millot, R.; Kloppmann, W.; Vengosh, A. New Tracers Identify Hydraulic Fracturing Fluids and Accidental Releases from Oil and Gas Operations. *Environ. Sci. Technol.* **2014**, *48* (21), 12552−12560.
(11) Kibbey, T. C. G.; Jabrzemski, R.; O'Carroll, D. M. Supervised machine learning for source allocation of per- and polyfluoroalkyl substances (PFAS) in environmental samples. *Chemosphere* **2020**, *252*, 126593.
(12) Allan, S. E.; Smith, B. W.; Anderson, K. A. Impact of the Deepwater Horizon Oil Spill on Bioavailable Polycyclic Aromatic Hydrocarbons in Gulf of Mexico Coastal Waters. *Environ. Sci. Technol.* **2012**, *46* (4), 2033−2039.
(13) 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. *Environmental Science & Technology Letters* **2016**, *3* (9), 316−321.
(14) 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* (18), 10317−10327.
(15) Stets, E. G.; Sprague, L. A.; Oelsner, G. P.; Johnson, H. M.; Murphy, J. C.; Ryberg, K.; Vecchia, A. V.; Zuellig, R. E.; Falcone, J. A.; Riskin, M. L. Landscape Drivers of Dynamic Change in Water Quality of U.S. Rivers. *Environ. Sci. Technol.* **2020**, *54* (7), 4336−4343.
(16) Macrae, M.; Jarvie, H.; Brouwer, R.; Gunn, G.; Reid, K.; Joosse, P.; King, K.; Kleinman, P.; Smith, D.; Williams, M.; Zwonitzer, M. One size does not fit all: Toward regional conservation practice guidance to reduce phosphorus loss risk in the Lake Erie watershed. *Journal of Environmental Quality* **2021**, *50* (3), 529−546.
(17) Rose, V. J.; Forney, W. M.; Norton, R. A.; Harrison, J. A. Catchment characteristics, water quality, and cyanobacterial blooms in Washington and Oregon Lakes. *Lake and Reservoir Management* **2019**, *35* (1), 51−63.
(18) Naumburg, E.; Ellsworth, D. S.; Katul, G. G. Modeling dynamic understory photosynthesis of contrasting species in ambient and elevated carbon dioxide. *Oecologia* **2001**, *126* (4), 487−499.
(19) Stoy, P. C.; Katul, G. G.; Siqueira, M. B. S.; Juang, J.-Y.; McCarthy, H. R.; Kim, H.-S.; Oishi, A. C.; Oren, R. Variability in net ecosystem exchange from hourly to inter-annual time scales at adjacent pine and hardwood forests: a wavelet analysis. *Tree Physiology* **2005**, *25* (7), 887−902.
(20) Flanagan, L. B.; Wever, L. A.; Carlson, P. J. Seasonal and interannual variation in carbon dioxide exchange and carbon balance in a northern temperate grassland. *Global Change Biology* **2002**, *8* (7), 599−615.
(21) Belovsky, G. E.; Stephens, D.; Perschon, C.; Birdsey, P.; Paul, D.; Naftz, D.; Baskin, R.; Larson, C.; Mellison, C.; Luft, J.; Mosley, R.; Mahon, H.; Van Leeuwen, J.; Allen, D. V. The Great Salt Lake Ecosystem (Utah, USA): long term data and a structural equation approach. *Ecosphere* **2011**, *2* (3), art33.
(22) Dávila-Santiago, E.; Shi, C.; Mahadwar, G.; Medeghini, B.; Insinga, L.; Hutchinson, R.; Good, S.; Jones, G. D. Machine Learning Applications for Chemical Fingerprinting and Environmental Source Tracking Using Non-target Chemical Data. *Environ. Sci. Technol.* **2022**, *56* (7), 4080−4090.
(23) Biasi, C.; Rusalimova, O.; Meyer, H.; Kaiser, C.; Wanek, W.; Barsukov, P.; Junger, H.; Richter, A. Temperature-dependent shift from labile to recalcitrant carbon sources of arctic heterotrophs. *Rapid Communications in Mass Spectrometry: An International Journal Devoted to the Rapid Dissemination of Up-to-the-Minute Research in Mass Spectrometry* **2005**, *19* (11), 1401−1408.
(24) Lapierre, J.-F.; Del Giorgio, P. Partial coupling and differential regulation of biologically and photochemically labile dissolved organic carbon across boreal aquatic networks. *Biogeosciences* **2014**, *11* (20), 5969−5985.
(25) Roebuck, J. A., Jr; Seidel, M.; Dittmar, T.; Jaffé, R. Controls of Land Use and the River Continuum Concept on Dissolved Organic Matter Composition in an Anthropogenically Disturbed Subtropical Watershed. *Environ. Sci. Technol.* **2020**, *54* (1), 195−206.
(26) Williams, C. J.; Yamashita, Y.; Wilson, H. F.; Jaffé, R.; Xenopoulos, M. A. Unraveling the role of land use and microbial activity in shaping dissolved organic matter characteristics in stream ecosystems. *Limnology and Oceanography* **2010**, *55* (3), 1159−1171.
(27) Kothawala, D. N.; Ji, X.; Laudon, H.; Ågren, A. M.; Futter, M. N.; Köhler, S. J.; Tranvik, L. J. The relative influence of land cover, hydrology, and in-stream processing on the composition of dissolved organic matter in boreal streams. *Journal of Geophysical Research: Biogeosciences* **2015**, *120* (8), 1491−1505.
(28) Sachse, A.; Henrion, R.; Gelbrecht, J.; Steinberg, C. E. W. Classification of dissolved organic carbon (DOC) in river systems: Influence of catchment characteristics and autochthonous processes. *Org. Geochem.* **2005**, *36* (6), 923−935.
(29) Jansson, M.; Bergström, A.-K.; Blomqvist, P.; Drakare, S. ALLOCHTHONOUS ORGANIC CARBON AND PHYTOPLANKTON/BACTERIOPLANKTON PRODUCTION RELATIONSHIPS IN LAKES. *Ecology* **2000**, *81* (11), 3250−3255.
(30) Taylor, G. H.; Bartlett, A. Climate zone 2, Willamette Valley. *The Climate of Oregon*; OSU Press, 1993.
(31) Homer, C. H.; Fry, J. A.; Barnes, C. A. The national land cover database. *US geological survey fact sheet* **2012**, *3020* (4), 1−4.
(32) URycki, D. R.; Good, S. P.; Crump, B. C.; Chadwick, J.; Jones, G. D. River Microbiome Composition Reflects Macroscale Climatic and Geomorphic Differences in Headwater Streams. *Frontiers in Water* **2020**, DOI: 10.3389/frwa.2020.574728.
(33) Ries, K. G, III; Guthrie, J. D.; Rea, A. H.; Steeves, P. A.; Stewart, D. W. StreamStats: A water resources web application. *US Geological Survey fact sheet* **2008**, 2327−6932.
(34) Xia, Y.; Mitchell, K.; Ek, M.; Sheffield, J.; Cosgrove, B.; Wood, E.; Luo, L.; Alonge, C.; Wei, H.; Meng, J. *NLDAS NOAH land surface model l4 hly 0.125× 0.125 degree v002*; NLDAS: Greenbelt, MD, 2012; Vol. *1025*.
(35) Apprill, A.; McNally, S.; Parsons, R.; Weber, L. Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton. *Aquatic Microbial Ecology* **2015**, *75* (2), 129−137.
(36) Kozich, J. J.; Westcott, S. L.; Baxter, N. T.; Highlander, S. K.; Schloss, P. D. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq

https://doi.org/10.1021/acs.est.2c08540
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

Illumina sequencing platform. *Applied and environmental microbiology* **2013**, *79* (17), 5112−5120.
(37) Bolyen, E.; Rideout, J. R.; Dillon, M. R.; Bokulich, N. A.; Abnet, C. C.; Al-Ghalith, G. A.; Alexander, H.; Alm, E. J.; Arumugam, M.; Asnicar, F. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nature biotechnology* **2019**, *37* (8), 852−857.
(38) McDonald, D.; Price, M. N.; Goodrich, J.; Nawrocki, E. P.; DeSantis, T. Z.; Probst, A.; Andersen, G. L.; Knight, R.; Hugenholtz, P. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *ISME journal* **2012**, *6* (3), 610−618.
(39) McCune, B.; Grace, J. B.; Urban, D. L. *Analysis of ecological communities*; MjM software design Gleneden Beach: OR, 2002; Vol. *28*.
(40) Clarke, K. R. Non-parametric multivariate analyses of changes in community structure. *Australian journal of ecology* **1993**, *18* (1), 117−143.
(41) Sinclair, L.; Ijaz, U. Z.; Jensen, L. J.; Coolen, M. J.; Gubry-Rangin, C.; Chroňáková, A.; Oulas, A.; Pavloudi, C.; Schnetzer, J.; Weimann, A. Seqenv: linking sequences to environments through text mining. *PeerJ.* **2016**, *4*, No. e2690.
(42) Borcard, D.; Legendre, P.; Drapeau, P. Partialling out the spatial component of ecological variation. *Ecology* **1992**, *73* (3), 1045−1055.
(43) Peres-Neto, P. R.; Legendre, P.; Dray, S.; Borcard, D. Variation partitioning of species data matrices: estimation and comparison of fractions. *Ecology* **2006**, *87* (10), 2614−2625.
(44) Pagano, T.; Bida, M.; Kenny, J. E. Trends in Levels of Allochthonous Dissolved Organic Carbon in Natural Water: A Review of Potential Mechanisms under a Changing Climate. *Water* **2014**, *6*, 2862−2897.
(45) Kile, D. E.; Wershaw, R. L.; Chiou, C. T. Correlation of Soil and Sediment Organic Matter Polarity to Aqueous Sorption of Nonionic Compounds. *Environ. Sci. Technol.* **1999**, *33* (12), 2053−2056.
(46) Schwarzenbach, R. P.; Westall, J. Transport of nonpolar organic compounds from surface water to groundwater. Laboratory sorption studies. *Environ. Sci. Technol.* **1981**, *15* (11), 1360−1367.
(47) Spencer, R. G. M.; Butler, K. D.; Aiken, G. R. Dissolved organic carbon and chromophoric dissolved organic matter properties of rivers in the USA. *Journal of Geophysical Research: Biogeosciences* **2012**, *117* (G3), n/a.
(48) Lodge, D. J.; McDowell, W. H.; McSwiney, C. P. The importance of nutrient pulses in tropical forests. *Trends in Ecology & Evolution* **1994**, *9* (10), 384−387.
(49) Phillips, R. P.; Erlitz, Y.; Bier, R.; Bernhardt, E. S. New approach for capturing soluble root exudates in forest soils. *Functional Ecology* **2008**, *22* (6), 990−999.
(50) Nakayama, M.; Tateno, R. Solar radiation strongly influences the quantity of forest tree root exudates. *Trees* **2018**, *32* (3), 871−879.
(51) Wuest, S. Seasonal Variation in Soil Organic Carbon. *Soil Science Society of America Journal* **2014**, *78* (4), 1442−1447.
(52) Hill, S. K.; Hale, R. L.; Grinath, J. B.; Folk, B. T.; Nielson, R.; Reinhardt, K. Looking beyond leaves: variation in nutrient leaching potential of seasonal litterfall among different species within an urban forest. *Urban Ecosystems* **2022**, *25* (4), 1097−1109.
(53) Dickenson, E. R.; Snyder, S. A.; Sedlak, D. L.; Drewes, J. E. Indicator compounds for assessment of wastewater effluent contributions to flow and water quality. *Water research* **2011**, *45* (3), 1199−1212.
(54) Tian, Z.; Zhao, H.; Peter, K. T.; Gonzalez, M.; Wetzel, J.; Wu, C.; Hu, X.; Prat, J.; Mudrock, E.; Hettinger, R.; Cortina, A. E.; Biswas, R. G.; Kock, F. V. C.; Soong, R.; Jenne, A.; Du, B.; Hou, F.; He, H.; Lundeen, R.; Gilbreath, A.; Sutton, R.; Scholz, N. L.; Davis, J. W.; Dodd, M. C.; Simpson, A.; McIntyre, J. K.; Kolodziej, E. P. A ubiquitous tire rubber-derived chemical induces acute mortality in coho salmon. *Science* **2021**, *371* (6525), 185−189.
(55) Willis, G. H.; McDowell, L. L. Pesticides in agricultural runoff and their effects on downstream water quality. *Environmental Toxicology and Chemistry: An International Journal* **1982**, *1* (4), 267−279.
(56) Burant, A.; Selbig, W.; Furlong, E. T.; Higgins, C. P. Trace organic contaminants in urban runoff: Associations with urban land-use. *Environ. Pollut.* **2018**, *242*, 2068−2077.
(57) Peter, K. T.; Lundin, J. I.; Wu, C.; Feist, B. E.; Tian, Z.; Cameron, J. R.; Scholz, N. L.; Kolodziej, E. P. Characterizing the Chemical Profile of Biological Decline in Stormwater-Impacted Urban Watersheds. *Environ. Sci. Technol.* **2022**, *56* (5), 3159−3169.
(58) Ohlendorf, H. M.; Kilness, A. W.; Simmons, J. L.; Stroud, R. K.; Hoffman, D. J.; Moore, J. F. Selenium toxicosis in wild aquatic birds. *Journal of Toxicology and Environmental Health* **1988**, *24* (1), 67−92.
(59) Hohrenk-Danzouma, L. L.; Vosough, M.; Merkus, V. I.; Drees, F.; Schmidt, T. C. Non-target Analysis and Chemometric Evaluation of a Passive Sampler Monitoring of Small Streams. *Environ. Sci. Technol.* **2022**, *56* (9), 5466−5477.
(60) Cloutier, D. D.; Alm, E. W.; McLellan, S. L. Influence of Land Use, Nutrients, and Geography on Microbial Communities and Fecal Indicator Abundance at Lake Michigan Beaches. *Appl. Environ. Microbiol.* **2015**, *81* (15), 4904−4913.
(61) Hu, L. t.; Bentler, P. M. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: a multidisciplinary journal* **1999**, *6* (1), 1−55.
(62) Jones, G. D.; Droz, B.; Greve, P.; Gottschalk, P.; Poffet, D.; McGrath, S. P.; Seneviratne, S. I.; Smith, P.; Winkel, L. H. Selenium deficiency risk predicted to increase under future climate change. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114* (11), 2848−2853.
(63) Girardi, C.; Greve, J.; Lamshöft, M.; Fetzer, I.; Miltner, A.; Schäffer, A.; Kästner, M. Biodegradation of ciprofloxacin in water and soil and its effects on the microbial communities. *Journal of Hazardous Materials* **2011**, *198*, 22−30.
(64) Zhalnina, K.; Louie, K. B.; Hao, Z.; Mansoori, N.; da Rocha, U. N.; Shi, S.; Cho, H.; Karaoz, U.; Loqué, D.; Bowen, B. P.; Firestone, M. K.; Northen, T. R.; Brodie, E. L. Dynamic root exudate chemistry and microbial substrate preferences drive patterns in rhizosphere microbial community assembly. *Nature Microbiology* **2018**, *3* (4), 470−480.
(65) Wilson, H. F.; Xenopoulos, M. A. Effects of agricultural land use on the composition of fluvial dissolved organic matter. *Nature Geoscience* **2009**, *2* (1), 37−41.
(66) Wilson, H. F.; Xenopoulos, M. A. DIEL CHANGES OF DISSOLVED ORGANIC MATTER IN STREAMS OF VARYING WATERSHED LAND USE. *River Research and Applications* **2013**, *29* (10), 1330−1339.
(67) Bianchi, T. S.; Filley, T.; Dria, K.; Hatcher, P. G. Temporal variability in sources of dissolved organic carbon in the lower Mississippi River. *Geochim. Cosmochim. Acta* **2004**, *68* (5), 959−967.
(68) Crump, B. C.; Hobbie, J. E. Synchrony and seasonality in bacterioplankton communities of two temperate rivers. *Limnology and Oceanography* **2005**, *50* (6), 1718−1729.
(69) Berggren, M.; Laudon, H.; Haei, M.; Ström, L.; Jansson, M. Efficient aquatic bacterial metabolism of dissolved low-molecular-weight compounds from terrestrial sources. *ISME Journal* **2010**, *4* (3), 408−416.
(70) Tanentzap, A. J.; Fitch, A.; Orland, C.; Emilson, E. J. S.; Yakimovich, K. M.; Osterholz, H.; Dittmar, T. Chemical and microbial diversity covary in fresh water to influence ecosystem functioning. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116* (49), 24689−24695.
(71) Staley, C.; Gould, T. J.; Wang, P.; Phillips, J.; Cotner, J. B.; Sadowsky, M. J. Bacterial community structure is indicative of chemical inputs in the Upper Mississippi River. *Frontiers in Microbiology* **2014**, *5*. DOI: 10.3389/fmicb.2014.00524

https://doi.org/10.1021/acs.est.2c08540
Environ. Sci. Technol. XXXX, XXX, XXX−XXX