Source water vulnerability to elevated total dissolved solids within a mixed-use Appalachian River basin

Eric R. Merriam¹, Michael P. Strager²*, J. Todd Petty³

¹ U.S. Army Corps of Engineers, Pittsburgh District, Pittsburgh, Pennsylvania, United States of America, ² Davis College of Agriculture, Natural Resources, and Design, West Virginia University, Morgantown, West Virginia, United States of America, ³ Forestry and Environmental Conservation, Clemson University, Clemson, South Carolina, United States of America

*mstrager@wvu.edu

Abstract

The upper Monongahela River basin in West Virginia represents a watershed wherein historic land use activities, including coal mining, conventional oil and gas development, and residential and urban development have contributed to elevated total dissolved solids (TDS) concentrations within surface and source waters with widespread exceedance of the 500 mg/L secondary drinking water standard for TDS. Our research was designed to characterize spatial variability in and sources of elevated and assess vulnerability of surface and source waters to additional TDS sources and loading. We compiled 3439 unique water quality sampling records throughout the study area over a ten-year period and applied boosted regression tree (BRT) analysis to model log[x]-transformed TDS as a function of landscape attributes across the 885 sampled NHD catchments. We predicted (i.e., elevated TDS and reduced dilution capacity) and observed (i.e., increased variability in TDS and temporary exceedances of 500 mg/L) elevated vulnerability of source waters, with vulnerability being largely driven by a few dominant contributors of TDS throughout the stream network. Notably, seven 12-digit hydrologic unit code (HUC) inflows to the Monongahela River and its vulnerable tributaries were predicted to require a combined 6,664,879 m³ of water (i.e., chemical footprint) to dilute TDS concentrations below 500 mg/L from July through September. Our results provide insights into how the mixed land uses containing extractive industries can impact water quality and demonstrates the importance of both spatial and temporal processes in controlling vulnerability of source waters to water pollution.

Introduction

Human activities on the landscape have contributed to the salinization—and in a broader sense, an increase in total dissolved solids (TDS)—of freshwater systems within the United States and across the globe [1]. Elevated TDS alters the structure (e.g., shifts in chemical and community composition) and processes (e.g., decomposition, nutrient cycling, and trophic
structure) of freshwater ecosystems, impacting the sustainability of goods and services they provide and the socioeconomic conditions of reliant communities [2–5]. Elevated TDS can reduce human health and wellbeing via degraded drinking water quality [6]. Constituents contributing to elevated TDS can mobilize toxic metals from soil, sediments, and infrastructure and cause or exacerbate health conditions such as hypertension and cardiac disease [6–8]. Elevated TDS can also lead to corrosion and scaling within water distribution infrastructure, increasing treatment and maintenance costs [6].

The U.S. Environmental Protection Agency (US EPA) has established a non-enforceable 500 mg/L secondary drinking water standard for TDS [7]. However, the myriad of environmental and socioeconomic risks associated with elevated TDS, in conjunction with continued and widespread increases in TDS concentrations within freshwater systems, have led to recent calls for more comprehensive regulation and management of salts and other dissolved constituents within freshwaters [1, 9, 10]. Central to this effort is the need to characterize and quantify the contribution of various anthropogenic sources to elevated TDS and use this information to inform and prioritize management and regulatory actions [10]. This is particularly true for watersheds wherein historic and contemporary land use activities have resulted in high vulnerability of source waters to additional TDS loading [10].

The upper Monongahela River basin in West Virginia represents one such watershed wherein historic land use activities, including coal mining, conventional oil and gas development, and residential and urban development have contributed to elevated TDS concentrations within surface and source waters, increasing their vulnerability to additional TDS sources and loading [11–13]. The upper Monongahela River basin is currently experiencing an expansion of unconventional oil and gas development, which can increase TDS concentrations within surface waters downstream of extraction and disposal sites [14–16]. The goal of this research was to facilitate management and regulation of existing and potential future sources of TDS within the upper Monongahela River basin. Our specific objectives were to: (1) characterize spatial variability in and sources of elevated TDS throughout the upper Monongahela River basin, and (2) assess vulnerability of surface and source waters to additional TDS sources and loading.

Study area

The upper Monongahela River basin is part of the Appalachian Plateau physiographic province and drains approximately 10,720 km² within West Virginia and Pennsylvania and is comprised of the Cheat River, Tygart River, West Fork, and Monongahela 8-digit hydrologic unit code (HUC) watersheds (Fig 1). The topography is rugged with a highly dissected plateau with deeply eroded stream valleys [17]. Elevation values range from 215 to 1,482 meters with varying degrees of sloped land eroded by the stream valleys. Geology consists of sedimentary rocks (sandstone, shale, coal, and limestone) of Devonian, Mississippian, Pennsylvanian, and Permian age [17]. Soils are generally classified for this area as either being mostly shallow, weakly developed, poorly drained and have low fertility due to the higher slopes (the Monongahela-Zoar-Allegheny association) or well-drained and fertile on gentler slopes over unconsolidated sediments (Westmoreland-Culleoka-Clarksburgh association) [17].

The Monongahela River forms at the confluence of the Tygart River and West Fork in Fairmont, West Virginia, and the Cheat River enters the Monongahela River in Point Marion, Pennsylvania. The study area is predominately forested (78%); however, the drainage network is influenced by pre- and post-Surface Mine Control and Reclamation Act (SMCRA) surface mining (3%) and residential and urban development (2%). The Monongahela River Basin is considered to be a humid continental climate due to its mid-latitude. The average annual
Fig 1. Location of sites used to model spatial \((n = 885)\) patterns in total dissolved solids (TDS) within 8-digit hydrologic unit code (HUC) watersheds comprising the upper Monongahela River basin within West Virginia and Pennsylvania. Sites are separated into those where TDS was measured and those where TDS was estimated as a function of specific conductivity. Location of drinking water intakes \((n = 32)\) are also shown. Spatially mapped data for the state boundaries, hydrologic unit code watersheds, rivers, reservoirs, and drinking water intakes was obtained from the WV GIS Tech Center (wvgis.wvu.edu).

https://doi.org/10.1371/journal.pwat.0000035.g001
temperatures are lowest in the higher elevation mountainous areas of the south and east. Most of the weather patterns arrive from the westerly direction with northeasterly hurricanes possible from June to November. Average annual temperatures range from 5 to 11 C while precipitation is rather evenly distributed across the Basin at 104 cm a year.

The stream network also drains 905 deep mine National Pollution Discharge Elimination System (NPDES) permits, as well as 22,249 conventional (i.e., vertical wells drilled into formations not requiring hydraulic fracturing) and 743 unconventional (i.e., horizontal wells drilled into formations requiring hydraulic fracturing) oil and gas wells. A full description of land cover sources and methods is provided in the subsequent section. Flows within the upper Monongahela River drainage are regulated by two U.S. Army Corps of Engineers reservoirs—Stonewall Jackson Lake on the West Fork River and Tygart Lake on the Tygart Valley River [18] Fig 1. Both reservoirs were authorized for flood protection and water supply. Water releases from both reservoirs during low-flow periods help maintain commercial navigation and improve water quality for domestic (e.g., drinking water) and ecological (e.g., fish and wildlife habitat) purposes. Flows within the Cheat River are regulated by a private hydroelectric dam (Fig 1). There are 32 surface water intakes used for public water supply throughout the upper Monongahela River watershed (Fig 1) [19].

Methods

Water quality data
We compiled 3439 unique water quality sampling records collected by the WV DEP throughout the study area during the summers (7/1–9/30) of 2006 through 2016. We took several steps to validate the spatial water quality dataset. We used sampling location notes to first remove samples taken from lentic systems, as well as samples taken from discharges, seeps, and springs. Remaining collection records were spatially linked to National Hydrography Dataset-Plus Version 2 catchments [20]. We visually checked the accuracy of NHD catchment assignments for all remaining samples. We removed sites for which correct assignment could not be verified and/or the sampling location was not representative of upstream land use. A total of 2692 samples taken from 889 NHD catchments [Cheat River (CR; n = 212), Tygart Valley River (TV; n = 308), West Fork (WF; n = 216), and Monongahela River (MR; n = 153)] were retained for further analysis.

TDS was measured within 612 NHD catchments according to the WV DEP sampling records. The number of TDS measurements within the CR watershed (n = 48) was considerably less than for the TV (n = 221), WF (n = 211), and MR (n = 132) watersheds (Fig 1). Given this spatial unevenness, we used linear regression to estimate TDS as a function of in situ specific conductance for all 889 sampled NHD catchments.

Landscape data
We used the NHD Plus V2 Catchment Attribute Allocation and Accumulation Tool (CA3TV2; [20]) in conjunction with ArcGIS ArcMap (version 10.2; Environmental Systems Research Institute, Redlands, California) to calculate cumulative measures of each landscape attribute for all NHD catchments within the study area (n = 5302). We classified publicly available high resolution National Agricultural Imagery Program (NAIP) aerial imagery to characterize land use and cover within the West Virginia and Pennsylvania portions of the study [21]. Base land cover classifications across the entire study area included forest, grassland, cropland, barren, open water, development, and impervious. Roads were distinguished from other impervious cover.
We used spatial data created by the WV DEP (https://tagis.dep.wv.gov/site/GISData) and Pennsylvania Department of Environmental Protection (PA DEP; https://newdata-padep-1.opendata.arcgis.com) to further characterize mining and oil and gas development throughout the study area. We leveraged known abandoned mine land extents (i.e., pre-Surface Mine Control and Reclamation Act; US Congress 1977) and surface mine permit boundaries for permits mined after the implementation of SMCRA to differentiate mining-related forest, grassland, and barren land. Mining classes were summed to obtain measures of the total percentage of upstream land area mined prior (pre-SMCRA) and subsequent (post-SMCRA) to implementation of SMCRA. We obtained underground mine NPDES permit and conventional oil and gas well locations from the WV DEP and PA DEP and used these data to calculate the cumulative density of underground mine NPDES permits (no./km$^2$) and conventional oil and gas wells (no./km$^2$). In order to account for the rapid expansion of unconventional oil and gas development within the study area, we attributed each unconventional oil and gas well with the date of first disturbance as identified from aerial imagery and used these data to create an annual time series of upstream unconventional oil and gas well density (no./km$^2$) for each catchment [22]. Each water quality sample was matched to the appropriate annual unconventional oil and gas well density. We also classified unconventional oil and gas well pad disturbance and incorporated these extents into the base land cover dataset [22]. The full list of landscape attributes considered during model development can be found in the supporting information (S1 Table).

Modeling spatial variability in TDS

We used boosted regression tree (BRT) analysis to model log[$x$]-transformed TDS as a function of landscape attributes across the 885 sampled NHD catchments with the full suite of landscape attributes. BRT is an extension of classification and regression tree analysis, whereby many simple models are iteratively fit to random subsets of the data (bag fraction) and combined to better estimate the response [23]. Successive trees focus on unexplained variance, resulting in improved model accuracy and enhanced predictive performance [24]. BRT is particularly well-suited for modeling environmental data from landscape datasets having large numbers of predictors with complex covariance structures [25]. We retained only the most recent sample taken from each catchment for model development.

We removed statistically redundant predictors ($\rho > 0.70$) (S2 Table). The final predictor set consisted of 12 variables representing general land cover (upstream forest, barren, water, and wetlands) and use (impervious cover, pre-SMCRA mining, post-SMCRA mining, deep mining, and unconventional and conventional oil and gas development), as well as watershed characteristics (basin area and 8-digit HUC watershed) (Table 1). We created an initial global model using all 12 variables. The model was optimized by varying tree complexity (i.e., number of splits within each tree and degree of interaction) and learning rate (i.e., contribution of successive trees to the growing model) to minimize predictive error sensu [24]. We set tree complexity to two because previous research with this region indicates two-way interactions are important when predicting TDS from landscape characteristics within this region [26], and performance did not improve with greater interaction depth. We set learning rate to 0.01, which maximized deviance reduction without overfitting (i.e., increase in predictive deviance with successive trees). Bag fraction was set to 0.5—meaning 50% of the data were drawn at random during each iteration—introducing stochasticity and maximizing predictive performance [24]. We used functions in R package ‘dismo’ to develop BRT models [27].

We assessed predictive performance of the final model by calculating model and mean cross-validated (CV) deviance explained. We quantified relative influence of predictor variables and constructed partial dependency plots to show effects [24]. The relative strength of all
two-way interactions was quantified using the function gbm.interaction in R package 'dismo', which estimates residual variance of a linear model fitted to predictions from the joint effect of two variables [27]. Larger residual variance indicates stronger interactions [24]. We conservatively report only the top-ranked interactions.

Surface and source water vulnerability

Vulnerability to water pollution is a function of current contaminant levels (i.e., concentrations) and the assimilative capacity of receiving system [28]. We used the final BRT model to predict summer TDS concentrations within all un-sampled stream segments (n = 5302) and quantified the number of stream segments with predicted TDS exceeding the 500 mg/L secondary drinking water standard [7]. We summarized predicted TDS for 12-digit HUC watershed outflows (n = 109) and streams with source water intakes used for public water supply (n = 32; [19]. Predicted TDS concentrations were mapped (ArcMap version 10.7.1; Environmental Systems Research Institute, Redlands, California) to show spatial patterns.

We converted predicted TDS concentrations into a chemical pollution footprint (subsequently referred to as chemical footprint; ChF) to quantify assimilative capacity for all streams segments [29]. ChF was calculated via the following equation:

$$ChF = Q - \left( \frac{Q \times OPC}{TPC} \right)$$

where Q (m$^3$) is water yield over the three-month summer period (July–September) and OPC (kg/m$^3$) and TPC (kg/m$^3$) are the observed and threshold (500 mg/L) pollutant concentrations, respectively. We leveraged extended unit runoff method mean monthly summer (July–September) stream flows contained within the NHD-Plus V2 to obtain estimates of Q for all streams [20]. ChF reflects the volume of water available to dilute additional TDS loading (ChF >0) or required to dilute existing TDS loads to allowable levels (i.e., 500 mg/L; ChF <0). We summarized ChF for 12-digit HUC watershed outflows (n = 109) and streams with drinking water intakes (n = 32).

Surface and source water vulnerability was assessed by examining predicted TDS and ChF across 12-digit HUC outflows and streams with source water intakes used for public water supply [19]. We compared vulnerability as characterized by predicted TDS and ChF to observed TDS (i.e., maximum and standard deviation) across the outflows of 39 12-digit HUC watersheds and 4 stream segments associated with drinking water intakes with historic TDS.
samples. Sites with observed TDS values were classified as either ‘meeting’ drinking water standards (i.e., both minimum and maximum TDS <500mg/L), ‘not meeting’ drinking water standards (i.e., both minimum and maximum observed TDS >500 mg/L), and ‘vulnerable’ to changes in TDS (i.e., minimum TDS <500 mg/L and maximum TDS ≥500 mg/L). To further characterize the role temporal variability plays in determining vulnerability, we ran a one-way analysis of variance (ANOVA) and Tukey post-hoc comparison to test for significant differences in standard deviation of observed TDS across sites designated as ‘impaired’ and those with ‘high vulnerability’ and ‘low vulnerability’.

Results
Modeling spatial variability in TDS
Two outliers were identified and removed from the linear regression predicting TDS as a function of in situ specific conductance (log[x] = log[y]×1.02–0.50; $R^2$ = 0.97; $p < 0.001$). Given the strong relationship between specific conductance and TDS, we felt comfortable modeling estimated TDS as a function of landscape attributes. Boosted regression analysis was run across 884 sites for which predicted TDS and the full suite of landscape attributes were available.

The final BRT model incorporated 3700 trees and made use of 10 variables to model TDS with a high degree of accuracy (model deviance explained = 0.84 and CV deviance explained = 0.70; Table 1). Pre-SMCRA mining (RI = 23.1%; Fig 2A) and post-SMCRA mining

![Functional response curves](https://doi.org/10.1371/journal.pwat.0000035.g002)

Fig 2. Functional response curves (solid lines) for select landscape predictor variables retained in the final model predicting log[x]-transformed TDS within the upper Monongahela River basin, including pre-Surface Mine Control and Reclamation Act (SMCRA) mining (A), 8-digit hydrologic unit code (HUC) watershed (B), forested land cover (C), post-SMCR A mining (D), impervious land cover (E), and deep mine NPDES permit density (F). Dotted lines represent the smoothed form of the fitted functions (Elith et al. 2008).
RI = 7.5%; Fig 2D) were the dominant anthropogenic predictors of TDS throughout the Monongahela River watershed, followed by impervious cover (RI = 7.5%; Fig 2E), deep mine density (RI = 4.5%; Fig 2F), and conventional oil and gas well density (RI = 4.4%)—all of which exhibited consistent positive relationships with TDS (Table 1). Forest exhibited a strong negative relationship with TDS (RI = 14.9%; Fig 2C). Forest interacted with conventional oil and gas well density, with gas wells having an incrementally greater effect on TDS in streams with greater forested area (relative interaction strength = 4.4; Fig 3A). We observed an interaction between basin area (RI = 5.5%) and percentage of the land area covered in surface water (RI = 7.2%) whereby TDS decreases strongly with increasing basin area and surface water (relative interaction strength = 1.1; Fig 3B). Watershed (i.e., CR, TV, WF, or MR) was also a dominant predictor of TDS (RI = 20.2%) and exhibited a strong interaction with pre-SMCRA mining (relative interaction strength = 4.4.; Fig 4A), forest (2.05; Fig 4B), impervious cover (2.00; Fig 4C), and deep mine density (1.6; Fig 4D) whereby equivalent levels of each were associated with higher TDS in the MR and WF watersheds as compared to the CR and TV watersheds.

**Surface and source water vulnerability**

Predicted TDS and ChF ranged from 10 to 4417 mg/L and -13,232,602 to 295,529,395 m$^3$ respectively across all stream segments ($n = 5302$) within the upper Monongahela River basin (Fig 5A and 5B). A total of 634 (12%) stream segments were predicted to exceed the 500 mg/L secondary drinking water standard throughout the study area. Predicted exceedances were more prevalent within the WF ($n = 366$; 34%) and MR (201; 32%) watersheds as compared to the TV (28; 1%) and CR (39; 2%) watersheds (Fig 5A and 5B). Predicted TDS and ChF ranged from 24 to 1389 mg/L and -8,908,290 to 292,639,911 m$^3$ respectively across 12-digit HUC.
watershed outflows (Fig 5C and 5D). The outflows of 14 12-digit HUC watersheds—nine in the WF and five in the MR watersheds—were predicted to exceed 500 mg/L. Seven 12-digit HUC outflows exceeding 500 mg/L drain directly to the WF and MR, with a combined ChF of -6,664,879 m$^3$. No 12-digit HUC watershed outflows were predicted to exceed 500 mg/L in the TV and CR watersheds. Predicted TDS and ChF ranged from 23 to 454 and 123,935 to 243,340,317 m$^3$ respectively across stream segments with drinking water intakes (Fig 5A and 5B).
Fig 5. Maps of predicted TDS and chemical footprint (ChF) for all stream segments (A and B, respectively; \( n = 5302 \)) and 12-digit hydrologic unit code (HUC) watershed outflows \( (n = 109) \) within the 8-digit HUC watersheds comprising the upper Monongahela River basin. Location of reservoirs and drinking water intakes are also shown. Spatially mapped data for the hydrologic unit code watersheds, rivers, reservoirs, and drinking water intakes was obtained from the WV GIS Tech Center (wvgis.wvu.edu).

https://doi.org/10.1371/journal.pwat.0000035.g005
Numerous 12-digit HUC watershed outflows and streams with drinking water intakes were predicted to have elevated TDS concentrations (i.e., approaching the 500 mg/L) and reduced assimilative capacity (i.e., ChF near zero), indicating increased vulnerability to additional TDS loading (Fig 6A). There was strong agreement between predicted and observed vulnerability.

Fig 6. (A) Predicted TDS and chemical footprints (ChF) for all 12-digit HUC outflows \((n = 109)\) and stream segments with drinking water intakes \((n = 32)\) within the upper Monongahela River basin. The horizontal dashed line indicates the 500 mg/L secondary drinking water standard and the vertical dashed line indicates a ChF of zero (i.e., no capacity to assimilate additional TDS loading). The arrows indicate increasing vulnerability as characterized by elevated predicted TDS and decreasing ChF. (B) Predicted TDS and chemical footprints (ChF) for 12-digit HUC outflows \((n = 39)\) and stream segments with drinking water intakes \((n = 4)\) having observed TDS measures. Sites are characterized based on observed TDS as ’meeting’ drinking water standards (i.e., minimum and maximum TDS \(<500\text{ mg/L}\) ), ’not meeting’ drinking water standards (i.e., minimum and maximum observed TDS \(\geq 500\text{ mg/L}\) ), and ’vulnerable’ to changes in TDS (i.e., minimum TDS \(<500\text{ mg/L}\) and maximum TDS \(\geq 500\text{ mg/L}\) ). Symbol size reflects relative variability of observed values (standard deviation).

https://doi.org/10.1371/journal.pwat.0000035.g006
Ten 12-digit HUC watershed outflows were categorized as ‘not meeting’ drinking water standards based on observed TDS values (i.e., minimum and maximum observed TDS > 500 mg/L)—seven of which (70%) were predicted to have TDS > 500 mg/L and ChF < 0. Twenty-three 12-digit HUC outflows and three drinking water intakes were observed to be ‘meeting’ drinking water standards (i.e., minimum and maximum observed TDS < 500 mg/L)—all but one of which (96%) had predicted TDS < 400 mg/L and positive ChF (Fig 6B). Six 12-digit HUC outflows and one drinking water intake were categorized as ‘vulnerable’ based on observed TDS values (i.e., minimum TDS < 500 mg/L and maximum TDS > 500 mg/L)—all but two of which had predicted TDS > 400 mg/L and ChF close to zero. The remaining two ‘vulnerable’ streams were located along the WF mainstem and had higher variability in observed TDS than streams with similar predicted TDS and/or ChF observed to be ‘meeting’ drinking water standards (Fig 6B).

Discussion
Elevated TDS within source waters represents an important health, economic, and engineering issue [6]. Elevated TDS can cause corrosion and scaling within water distribution infrastructure, increasing treatment and maintenance requirements and costs [6]. Constituents contributing to elevated TDS within the Monongahela River basin [e.g., chloride [30] and sulfate [31] have been shown to mobilize toxic metals from aging infrastructure, posing a direct health risk [6, 13]. Moreover, elevated salt content can be a concern for people with and exacerbate certain health conditions such as hypertension and cardiovascular and kidney disease [6–8]. Mixed and pervasive historic and contemporary land use activities within this and other Appalachian basins have resulted in elevated TDS concentrations with complex chemical mixtures [13, 26]. Further study is warranted into how water distribution systems within this and other vulnerable Appalachian watersheds are potentially impacted by elevated TDS.

A limitation of this study is that it only considers source water vulnerability to a single water quality parameter—TDS. Although the secondary drinking water standard of 500 mg/L represents an important criteria to preserve drinking water quality and prevent impacts to distribution systems (e.g. corrosion), other chemical stressors associated with dominant land use activities within this watershed, including those associated with mining (e.g., aluminum, iron, selenium, sulphate), oil and gas development (e.g., barium, chloride), and urban/residential development (e.g., sodium, chloride, fecal coliform) are also likely contributing to source water vulnerability. Because elevated TDS is associated with each of these major land use activities, however, managing TDS will inherently and collectively help reduce vulnerabilities associated with individual and contributing chemical stressors across all land use activities. Additional research aimed at better understanding how individual chemical stressors—particularly those not contributing to elevated TDS—are contributing to source water vulnerability is critical within this and other watersheds throughout the region.

Our modeling approach enabled us to quantify and characterize how historic and contemporary land use activities control spatial variability in and vulnerability to TDS. Our results corroborate previous studies identifying other land disturbance such as mining and deforestation from UOG development as the dominant contributor to elevated TDS within the Monongahela River basin [11, 13, 32]. The effect of mining varied strongly among watersheds, with equivalent levels of mining resulting in greater TDS concentrations within the WF and MR watersheds as compared to the CR and TV watersheds—a result that could reflect geologic controls over the severity of aquatic impacts of mining [33] or remediation of major acid mine drainage inputs that has primarily occurred within the CR and TV watersheds [32, 34]. In addition, as shown by [32] for the Monongahela River Watershed, the impacts on headwater
streams are likely from land disturbance and deforestation associated with UOG development and not from the produced water.

Unconventional oil and gas development did not significantly contribute to spatial variability in TDS within the upper Monongahela River basin. Conventional oil and gas development, which is much more pervasive throughout the study area (i.e., >22,000 conventional and >700 unconventional wells), was retained in the final model. Continued expansion of unconventional oil and gas development could result in significant TDS contributions within this and other actively developing watersheds. It is important to note that we considered spatial attributes characterizing the extent of unconventional oil and gas development on the landscape (i.e., disturbance footprints and well density), which would capture increases in TDS associated with general landscape disturbance [14]. We did not consider other mechanisms through which unconventional oil and gas development can increase TDS concentrations. Ephemeral increases in TDS can occur via spillage of produced waters and as a result of water withdrawals that reduce downstream assimilative capacity [14, 16]. Elevated TDS has also been observed downstream of deep well injection disposal sites [15] and publicly owned wastewater treatment plant facilities that accept produced waters [35, 36]. Future efforts to characterize the effects of unconventional oil and gas development on spatial variation in TDS should incorporate—to the extent possible—known disposal practices and sites.

Large contributors of TDS have increased vulnerability of source waters within the WF and MR watersheds. Notably, 12-digit HUC inflows to the WF and MR mainstems were predicted to require a combined 6,664,879 m$^3$ of water to dilute their concentrations below 500 mg/L during the summer months (July–September). This reduction in assimilative capacity increases the vulnerability of the WF and MR. Indeed two 12-digit HUC watersheds on the WF mainstem were classified as ‘vulnerable’ based on observed TDS concentrations (i.e., minimum TDS <500 mg/L and maximum TDS >500 mg/L). Our results corroborate other studies documenting increased vulnerability of the WF mainstem to flow-mediated variability in TDS [13] and further suggest that numerous source waters throughout the upper Monongahela River basin are vulnerable to even minor increases in TDS loading. Vulnerability of the WF (Stonewall Jackson Reservoir) and MR (Stonewall Jackson and Tygart River reservoirs) mainstems is reduced by upstream reservoirs as evidenced by the interaction between basin area and upstream open water within the final model. We agree with previous calls for additional research into whether reservoir management could be altered to further reduce vulnerability of downstream source waters while maintaining all other authorized purposes (i.e., flood protection, water supply, fish and wildlife habitat) [13].

Climate change is expected to affect the vulnerability of source waters within this and other watersheds by modulating spatial and temporal variability in pollution concentrations [37, 38]. Precipitation is expected to increase within the upper Monongahela River basin throughout the 21st century; however, increases in precipitation will largely be offset by increases in evapotranspiration [39, 40]. Changes in precipitation are also expected to be strongly seasonal, with the smallest increases occurring during the growing season when increases in evapotranspiration are expected to be greatest [41]. These results suggest that the upper Monongahela River basin will experience more extreme high-flow events during the dormant season (late fall through early spring) and more extreme low-flow events during the growing season (late spring through early fall). More severe and prolonged periods of drought would negatively affect water quality within the upper Monongahela River basin, particularly across vulnerable source waters within the WF and MR basins. More extreme periods of low flow could also impact the ability of upstream reservoirs to augment downstream flow, resulting in reduced assimilative capacity within downstream source waters [13, 40]. Climate change could also alter the volume of mine drainage and the rate of filling and subsequent required pumping of
underground mine pools. An important avenue of continued research will be to characterize how climate change will alter vulnerability of source waters within this and other at-risk watersheds.

Conclusions

The goal of this research was to investigate existing and potential future sources of TDS within a watershed with mixed land uses that also have extractive industries. This was an important question since historic land use activities—predominantly mining, but also increases in imperviousness and extensive conventional (i.e., vertical wells not requiring hydraulic fracturing) oil and gas development—have resulted in widespread exceedance of the 500 mg/L secondary drinking water standard for TDS. Our work can facilitate the management and regulation of existing and future sources of TDS within the upper Monongahela River basin with three notable results.

First, historic land use activities—predominantly mining, but also increases in imperviousness and conventional oil and gas development—have resulted in widespread exceedance of the 500 mg/L secondary drinking water standard for TDS within the upper Monongahela River basin. Although no stream segments with drinking water intakes were predicted to exceed 500 mg/L, our models suggest several source waters, including the WF mainstem, are vulnerable (i.e., elevated TDS and reduced dilution capacity) to further additional TDS loading—results that were corroborated by large observed variability in TDS that result in temporary exceedances of 500 mg/L in several streams characterized as vulnerable.

Second, the results can be used to identify management priorities within this and other watersheds vulnerable to any number of chemical pollutants. ChF values can be used to strategically target streams for restoration actions (i.e., reducing in pollution loads) that maximally increase assimilative capacity within and reduce vulnerability of downstream source waters, as well as identify streams as protection priorities in an effort to maintain downstream assimilative capacity and resiliency in the face of continued land use development and more extreme flow-mediated variability under climate change [29, 42]. Our model and associated TDS and ChF predictions can also facilitate regulatory decisions by providing a framework for assessing whether proposed actions throughout the watershed will likely result in unacceptable impacts to downstream source waters. The results highlight the importance of programs designed to ensure mine reclamation and treatment of mine drainage. Prioritization of these programs will be increasingly important as declines in coal production and the resulting bankruptcy of mining companies reduce funding for and place additional strain on state and federal reclamation programs.

And lastly, our results demonstrate the importance of both spatial and temporal processes in controlling vulnerability of source waters to water pollution. Vulnerability to water pollution is a function of background pollution concentrations (i.e., predicted summer TDS concentration) and assimilative capacity (i.e., ChF)—both of which are controlled by spatial variability in land use and associated TDS loading [28]. Higher background TDS concentrations require less additional loading to result in exceedance of drinking water standards. At a given TDS concentration, smaller ChF values indicate reduced capacity (i.e., volume of water) to assimilate additional TDS sources and loads. Streams with elevated TDS and reduced ChF are also more vulnerable to exceedances resulting from temporal variability in TDS. Six 12-digit HUC outflows and one drinking water intake were classified as ‘vulnerable’ based on observed TDS (i.e., temporary exceedances of 500 mg/L). Most of these sites (71%) were characterized as vulnerable based on predicted TDS and ChF (i.e., predicted TDS >400 mg/L and low ChF), indicating that elevated TDS and reduced assimilative capacity have increased
vulnerability to variability in TDS. This result corroborates other studies within this [13] and other watersheds [43] noting the importance of temporal variability in modulating socio-economic and environmental impacts of elevated TDS.

Supporting information

S1 Fig. Functional response curves predicting log[x]-transformed TDS within the upper Monongahela River basin, including pre-Surface Mine Control and Reclamation Act (SMCRA) mining (A), 8-digit hydrologic unit code (HUC) watershed (B), forested land cover (C), post-SMCRA mining (D), impervious land cover (E), and deep mine NPDES permit density (F), basin area (G), barren land cover (H), deep mine density (I), conventional oil and gas well density (J). (solid lines) for select landscape predictor variables retained in the final model Dotted lines represent the smoothed form of the fitted functions (Elith et al. 2008).

S1 Table. Range (and mean) across model sites (n = 884) and all catchments within the study area for which all predictor variables were available (n = 5302) of the 17 predictor variables summarized as potential predictors of total dissolved solids within the upper Monongahela River basin. HUC = hydrologic unit code; SMCRA = Surface Mine Control and Reclamation Act; NPDES = National Pollution Discharge Elimination System.

S2 Table. Results of Spearman correlations run to identify and remove redundant (i.e., rho > 0.70) predictor variables. Variables retained in the final model are indicated. Only correlations with rho > 0.70 are shown.

Author Contributions

Conceptualization: Eric R. Merriam, J. Todd Petty.
Data curation: Eric R. Merriam, Michael P. Strager, J. Todd Petty.
Formal analysis: Eric R. Merriam, Michael P. Strager, J. Todd Petty.
Funding acquisition: Michael P. Strager, J. Todd Petty.
Software: Michael P. Strager.
Writing – original draft: Eric R. Merriam.
Writing – review & editing: Michael P. Strager.

References

1. Kaushal SS, Likens GE, Pace ML, Utz RM, Haq S, Gorman J, et al. 2018. Freshwater salinization syndrome on a continental scale. Proceedings of the National Academy of Sciences 115:E574–E583. https://doi.org/10.1073/pnas.1711234115 PMID: 29311318
2. Kaushal SS, Groffman PM, Likens GE, Belt KT, Stack WP, Kelly VR, et al. 2005 Increased salinization of fresh water in the northeastern United States. Proceedings of the National Academy of Sciences 102:13517–13520. https://doi.org/10.1073/pnas.0506414102 PMID: 16157871
3. Canedo-Arguelles M., Kefford BJ, Piscart C, Prat N, Schafer RB, Schulz CJ. 2013. Salinization of rivers: an urgent ecological issue. Environmental Pollution 173:157–167. https://doi.org/10.1016/j.envpol.2012.10.011 PMID: 23202646

4. Berger E, Fror O, Schafer RB. 2018. Salinity impacts on river ecosystem processes: a critical mini-review. Philosophical Transactions of the Royal Society B 374:20180010. https://doi.org/10.1098/rstb.2018.0010 PMID: 30509912

5. Castillo AM, Sharpe DMT, Chalambor CK, and De Leon LF. 2018. Exploring the effects of salinization on trophic diversity in freshwater ecosystems: a quantitative review. Hydrobiologia 807:1–17.

6. Kaushal SS. 2016. Increased salinization decreases safe drinking water. Environmental Science & Technology 50:2765–2766. https://doi.org/10.1021/acs.est.6b00679 PMID: 26903048

7. U.S. Environmental Protection Agency (US EPA). 2018. 2018 Edition of the Drinking Water Standards and Health Advisory Tables. EPA 822-F-18-001. Office of Water, U.S. Environmental Protection Agency, Washington, DC, March 2018.

8. Chakraborty R, Khan KM, Dibaba DT, Khan MA, Ahmed A, Islam MZ. 2019. Health implications of drinking water salinity in coastal areas of Bangladesh. International Journal of Environmental Research and Public Health 16:3746. https://doi.org/10.3390/ijerph16193746 PMID: 31590283

9. Canedo-Arguelles M, Hawkins CP, Kefford BJ, Schafer RB, Dyack BJ, Brueet S, et al. 2016. Saving freshwater from salts. Science 351:914–916.

10. Schuler MX, Canedo-Arguelles M, Hintz WD, Dyack B, Birk S, Relyea RA. 2018. Regulations are needed to protect freshwater ecosystems from salinization. Philosophical Transactions of the Royal Society of London B: Biological Sciences 374(1764):20180019. https://doi.org/10.1098/rstb.2018.0019 PMID: 30509918

11. Merovich GT, Petty JT, Strager MP, Fulton JB. 2013. Hierarchical classification of stream condition: a house-neighborhood framework for establishing conservation priorities in complex riverscapes. Freshwater Science 32:874–891.

12. Wilson JM, Wang Y, Van Briesen JM. 2014. Sources of high total dissolved solids to drinking water supply in southwestern Pennsylvania. Journal of Environmental Engineering 140:B4014003.

13. Merriam ER, Petty JT, O’Neal M, Ziemkiewicz PF. 2020. Flow-mediated vulnerability of source waters to elevated TDS in an Appalachian river basin. Water 12:384.

14. Entrekine S, Evans-White M, Johnson B, Hagenbuch E. 2011. Rapid expansion of natural gas development poses a threat to surface waters. Frontiers in Ecology and the Environment 9:503–511.

15. Akob DM, Mumford AC, Orem W, Engle MA, Klinges JG, Kent DB, et al. 2016. Environmental Science & Technology 50:5517–5525.

16. Maloney KO, Baruch-Mordo S, Patterson LA, Nicot JP, Entrekine SA, Fargione JE, et al. 2017. Unconventional oil and gas spils: materials, volumes, and risks to surface waters in four states of the U.S. Science of the Total Environment 581–582:369–377. https://doi.org/10.1016/j.scitotenv.2016.12.142 PMID: 28043701

17. U.S. Army Corps of Engineers (US ACE). 2012. Monongahela River Watershed Initial Watershed Assessment September 2011 (Revised February 2012). Pittsburgh District, Pittsburgh Pennsylvania. https://www.lrp.usace.army.mil/Portals/72/docs/HotProjects/signed%20IWA_final_revised%20FE12%20public%20comments%20incorporated.pdf

18. U.S. Army Corps of Engineers (US ACE). 2019. Stonewall Jackson Lake, West Fork River, West Virginia. Design Memorandum No. 1, General Design Memorandum, Pittsburgh District, Pittsburgh Pennsylvania, 1971. https://www.lrp.usace.army.mil/Portals/72/docs/HotProjects/SWJ%20Public%20Release%20info1.pdf (accessed on 30 December 2019).

19. Source Water Assessment and Wellhead Protection Program (SWAWPS). 2018. WV Surface Water Intakes, West Virginia Bureau of Public Health, Office of Environmental Health Services, Environmental Engineering Division, Charleston, WV.

20. McKay L, Bondelid T, Dewald T, Johnson J, Moore R, Rea A. 2019. NHDPplus Version 2: User Guide.

21. Maxwell AE, Strager MP, Warner TA, Ramezan CA, Morgan AN, Pauley CE. 2019. Large-Area, High Spatial Resolution Land Cover Mapping using Random Forests, GEOBIA, and NAIP Orthophotography: Findings and Recommendations. Remote Sensing. Vol. 11, 1409, 1–27.

22. Harris KJ. "Disturbance Related to Unconventional Oil and Gas Development in the Appalachian Basin". (2020). Graduate Theses, Dissertations, and Problem Reports. 7548. https://researchrepository.wvu.edu/etd/7548. https://doi.org/10.33915/etd.7548

23. De’ath G. 2007. Boosted trees for ecological modeling and prediction. Ecology 88:243–251. https://doi.org/10.1890/0012-9658(2007)88[243:btfe2a]2.0.co;2 PMID: 17489472

24. Ellith J, Leathwick R, Hastie T. 2008. A working guide to boosted regression trees. Journal of Animal Ecology 77:802–813. https://doi.org/10.1111/j.1365-2656.2008.01390.x PMID: 18397250
25. Carlisle DM, Falcone J, Meador MR. 2009. Predicting the biological condition of streams: use of geospatial indicators of natural and anthropogenic characteristics of watersheds. Environmental Monitoring and Assessment 151:143–160. https://doi.org/10.1007/s10661-008-0256-z PMID: 18493861

26. Merriam ER, Petty JT, Strager MP, Maxwell AE, Ziemkiewicz PF. 2015. Complex contaminant mixtures in multistressor Appalachian riverscapes. Environmental Toxicology and Chemistry 34:2603–2610. https://doi.org/10.1002/etc.3101 PMID: 26053694

27. Hijmans RJ, Phillips S, Leathwick J, Elith J. 2017. dismo: species distribution modeling. R package version 1.1–4. https://CRAN.R-project.org/package=dismo

28. Tasdighi A, Arabi M, Osmond DL. 2017. The relationship between land use and vulnerability to nitrogen and phosphorus pollution in an urban watershed. Journal of Environmental Quality 46:113–122. https://doi.org/10.2134/jeq2016.06.0239 PMID: 28177402

29. Zip MC, Posthumus L, van de Meent D. 2014. Definition and applications of a versatile chemical pollution footprint methodology. Environmental Science & Technology 48:10588–10597. https://doi.org/10.1021/es500629f PMID: 25111657

30. Stets EG, Lee CJ, Lytle DA, Schock MR. 2018. Increasing chloride in rivers of the conterminous U.S. and linkages to potential corrosivity and lead action level exceedances in drinking water. Science of the Total Environment 613–614:1498–1509. https://doi.org/10.1016/j.scitotenv.2017.07.119 PMID: 28797521

31. Sun H, Shi B, Yang F, Want D. 2017. Effects of sulfate on heavy metal release from iron corrosion scales in drinking water distribution system. Water Research 114:69–77. https://doi.org/10.1016/j.watres.2017.02.021 PMID: 28262515

32. Michaels R, Eliason K, Kuzniar T, Petty JT, Strager MP, Ziemkiewicz PF, et al. 2022. Microbial communities reveal impacts of unconventional oil and gas development on headwater streams. Water Research 212:11807s. https://doi.org/10.1016/j.watres.2022.118073 PMID: 35091219

33. Petty JT, Fulton JB, Strager MP, Merovich GT, Stiles J, Ziemkiewicz P. 2010. Landscape indicators and thresholds of stream ecological impairment in an intensively mined Appalachian watershed. Journal of the North American Benthological Society 29:1292–1309.

34. Merovich GT, Stiles JM, Petty JT, Ziemkiewicz PF, Fulton JB. 2007. Water chemistry-based classification of streams and implications or restoring mined Appalachian watersheds. Environmental Toxicology and Chemistry 26:1361–1369. https://doi.org/10.1897/06-424r.1 PMID: 17655675

35. Wilson JM, Van Briesen JM. 2013. Source water changes and energy extraction activities in the Monongahela River, 2009–2012.

36. Wilson JM, Van Briesen JM. 2012. Oil and gas produced water management and surface drinking water sources in Pennsylvania. Environmental Practice 14:288–300.

37. Delpla I, Jung AV, Baures E, Clement M, Thomas O. 2009. Impacts of climate change on surface water quality in relation to drinking water production. Environment International 35:1225–1233. https://doi.org/10.1016/j.envint.2009.07.001 PMID: 19640587

38. Whitehead PG, Wilby RL, Battarbee RW, Keman M, Wade AJ. 2009. A review of the potential impacts of climate change on surface water quality. Hydrological Sciences Journal 54:101–123.

39. Gaertner BA, Zegre N, Warner T, Fernandez R, He Y, Merriam ER. 2019. Climate, forest growing season, and evapotranspiration changes in the central Appalachian Mountains, USA. Science of the Total Environment 650:1371–1381. https://doi.org/10.1016/j.scitotenv.2018.09.129 PMID: 30308824

40. Gaertner B, Fernandez R, Zegre N. 2020. Twenty-first century streamflow and climate change in forest catchments of the Central Appalachian Mountains Region, US. Water 12:453.

41. Fernandez R, Zegre N. 2019. Seasonal changes in water and energy balances over the Appalachian Region and beyond throughout the twenty-first century. Journal of Applied Meteorology and Climate 58:1079–1102.

42. Monfared SAH, Darmian MD, Snyder SA, Azizyan G, Pirzadeh B, Moghaddam MA. 2017. Water quality planning in rivers: assimilative capacity and dilution flow. Bulletin of Environmental Contamination and Toxicology 99, 531–541. https://doi.org/10.1007/s00128-017-2182-7 PMID: 28951956

43. Timpano AJ, Schoenholtz SH, Soucek DJ, Zipper CE. 2018. Benthic macroinvertebrate community response to salinization in headwater streams in Appalachia USA over multiple years. Ecological Indicators 91:645–656.