Inorganic constituents of conductivity in five central Appalachian watersheds with mixed source-driven pollutants

N.A. Cook a,*, L.H. Krometis a, E.A. Sarver b, J. Huang c

a Seitz Hall, RM 200, Virginia Tech, 155 AG Quad Lane, Blacksburg, VA 24061, United States
b Holden Hall, RM 108, Virginia Tech, Blacksburg, VA 24061-0239, United States
c Cheatham Hall, RM 105, Virginia Tech, 310 West Campus Drive, Blacksburg, VA 24061-0321, United States

ARTICLE INFO

Article history:
Received 7 January 2015
Received in revised form 26 March 2015
Accepted 22 April 2015
Available online 16 May 2015

Keywords:
Aquatic geochemistry
Alkaline mine drainage
Conductivity
Untreated household wastewater
Watershed science

ABSTRACT

Resource extraction activities (e.g., coal mining) are both a dominant economic driver and major ecological concern in the resource-rich region of Central Appalachia in the United States. Additionally, communities near watershed headwaters often lack effective wastewater treatment options and discharge untreated household wastes (UHW) directly into adjacent streams. The present study analyzes data collected from clustered watersheds on the Virginia–Kentucky border, which include sites influenced by discharges from active surface mines, a legacy underground mine, UHW, and minimally impacted (reference) sites. Over two years, in-situ measurements of conductivity were recorded, and grab samples were lab-analyzed for inorganic ion concentrations via ICP-MS. Patterns of detected ions were related to pollution types using a suite of multivariate statistical methods. Concentrations of inorganic ions identified the different pollution types, while permutational MANOVA testing (p = .001) to control for pseudoreplication (i.e., multiple sites along the same stream) confirmed that the groups were significantly different from each other. Notable associations were elevated P and UHW, Mn/Si/HCO3− and underground mining, and Ni, Ca, K, SO42−, and Se with surface mining. Given that mechanisms driving ecological impairments are likely related to specific contributors to conductivity, further efforts to examine relationships between particular ionic concentrations and macroinvertebrates will be useful in watershed remediation efforts in the Appalachian region.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

The Central Appalachian region of the United States consists of parts of Virginia, Kentucky, West Virginia and Pennsylvania. The region is generally synonymous with coal mining, given a lengthy history of resource extraction. In addition, it is characterized by mountainous terrain and related intricate hydrology, and significant biodiversity including high concentrations of unique plant and animal species (Bernhardt and Palmer, 2011). Juxtaposed against the significant environmental resources of the region is longstanding, widespread socio-economic distress as indicated by comparisons of numerous metrics to national averages (e.g., income, education, human health, etc.) (ARC, 2011; US Energy Information Administration, 2012). The complex interrelationships between environmental, economic, and social factors in Central Appalachia are evident in a host of regional issues, including many of those associated with water quality.

A large number of existing water quality impairments in the region have been associated with current and historic mining activities, which can contribute acidity, metals, sediment, or dissolved solids to streams via surface runoff and/or underground discharge points. The reaction of acids with local formations results in the release of ions common to that geology. Common ions associated with acid mine drainage (AMD) are Fe, Cu, Al, Co, and Ni, while K and HCO3− are generally associated with alkaline mine drainage (AlkMD). Both low and high pH discharges from mining are associated with elevated levels of Ca, Mg, Mn, and SO42− (Merovich et al., 2007; Lindberg et al., 2011; Bernhardt et al., 2012). Several studies have identified these inorganic constituents from specific sources including sediment ponds (Miguel-Chinchilla et al., 2014), stockpile runoff (Ibeanusi et al., 2003), and underground mine discharges of varying chemistries (Cravotta, 2008), as well as interactions between acid mine drainage (AMD) and stream sediments (Merovich et al., 2007).

Abbreviations: AMD, acid mine drainage; AlkMD, alkaline mine drainage; CA, cluster analysis; GIS, geographical information systems; ICP-MS, inductively-coupled plasma with mass spectrometry; PERMANOVA, permutational multiple analysis of variance; PEA, principle component analysis; UHW, untreated household wastewater; USEPA, United States Environmental Protection Agency.
* Corresponding author. Tel.: +1 828 291 8004.

http://dx.doi.org/10.1016/j.ecoleng.2015.04.090
0925-8374/ © 2015 Elsevier B.V. All rights reserved.
Collectively, discharges of inorganic ions contribute to measures of specific conductance (conductivity) in downstream waters. Given several recent studies that link elevated measures of conductivity to losses in benthic macroinvertebrate diversity and abundance (Schmidt et al., 2002; Sloane and Norris, 2003; Clements, 2004; Freund and Petty, 2007; Merricks et al., 2007; Pond et al., 2008; Pond, 2010; Cormier et al., 2013), conductivity is increasingly considered indicative of conditions affecting aquatic ecological health. Since these macroinvertebrates generally comprise the lower levels of the food chain in freshwater systems, such losses are considered a violation of the biological integrity standards under the Clean Water Act; stream condition indices (SCIs), which are used by many states to assess in-stream ecological health, are largely based on collective quantification of these organisms (Barbour et al., 1999). Ultimately, remediation efforts, including the development of total maximum daily loads (TMDLs) in accordance with the US Clean Water Act focus on identifying specific system stressors related to benthic macroinvertebrate loss; however, the specific mechanisms responsible for species loss are generally complex and not well understood (Benham et al., 2008; Keller and Cavallaro, 2008).

Oftentimes, field-scale studies focused solely on observed conductivity measurements do not account for additional potential confounding factors that may co-exist such as habitat loss, development in the riparian corridor, the presence of other pollutant sources, or the actual constituents of the conductivity observed (Schmidt et al., 2002; Sloane and Norris, 2003; Clements, 2004; Freund and Petty, 2007; Merricks et al., 2007; Pond et al., 2008; Pond, 2010; Cormier et al., 2013; Ziemkiewicz and Northway, 2013). Because of the limited data available, results from these field observations are not always readily comparable to studies attempting to identify maximum permissible levels of conductivity in laboratory mesocosms (Kennedy et al., 2004, 2005; Echols et al., 2010; Elphick et al., 2011a,b). Some attempts have been made to bridge these laboratory studies with field observations in order to explain the relatively high survivability of laboratory organisms at elevated conductivities (~1000 μS/cm) versus obvious species loss at lower levels (~300 μS/cm) in the field (USEPA 2011; Yeager-Armstead et al., 2013). The discontinuous nature of conductivity effects observed in the field and lab studies indicates that setting an effective conductivity benchmark may not be a simple task.

Although mining influences are perhaps the most readily identifiable sources of potential surface water contamination in this region, additional chronic issues related to economic distress may also contribute to losses in biological integrity. Across the more rural areas of the Central Appalachian coalfields, former “coal camp” communities often do not have access to wastewater treatment systems, and the local soils are inappropriate for traditional on-site septic systems. Consequently, the practice of directly discharging untreated household wastewater (UHW) comprised of both sewage and common household chemicals to nearby streams via “straight pipes” is common. Though this practice likely presents public health and ecological health risks, the issue is seldom addressed due to insufficient regulatory and financial resources (Tarricone 1989; Glasmeier and Farrigan, 2003; Gasteyer and Vaswani, 2004). No prior work has been published that describes the impacts of UHW on water quality or stream health in Central Appalachia, although some efforts are underway to describe mixed land uses (Merriam et al., 2011, 2013).

In an effort to study the relative effects of local mining activities and UHW discharges, 46 sites located within a cluster of five small watersheds were monitored monthly for two full years (25 sampling trips). Target watersheds were carefully selected to ensure similar underlying geologies. In-situ monthly monitoring of temperature, dissolved oxygen, pH, and conductivity was paired with grab samples analyzed for a suite of inorganic constituents. Subsequent statistical analyses aimed to: (1) identify particular inorganics that drive conductivity increases in this region; and (2) determine whether patterns of inorganic concentrations are distinct between primary known upland sources of surface water contamination.
2. Methodology

2.1. Study area

Preliminary GIS analysis and field reconnaissance, including meetings with local health agencies and sanitary surveys, were used to select study streams in watersheds of similar underlying geology but distinct primary pollution sources along the border of Virginia and Kentucky. The five study streams are all headwater streams; three are located in the Powell River basin in Wise County, Virginia, and two are located in the Cumberland River Basin in Harlan County, Kentucky. All streams are located in the Cumberland Plateau physiographic region, and are geologically dominated by Pennsylvania-age sandstones, shales, and coal formations. Land cover is primarily mixed-mesophytic hardwood forest as is typical of Central Appalachian systems, though the extent of forest cover varies with the level of mining activity (50% to >80%). The proportions of land classified as barren and/or grassland also vary with mining activity, from 19 to 21% in the watersheds with significant current surface mining footprints to as low as <1% in non-mining watersheds. Developed area in all watersheds is less than 10%, with communities clustered in the riparian corridors as is common for the steep, mountainous terrain typical of the region, which is dissected by small hollows. Land use is described here for illustrative purposes only, but was not formally used in the subsequent statistical analyses.

Target study streams included: Yocum Creek (Lower Clover Fork watershed, 8 sites) and Looney Creek (Looney Creek-Poor Fork watershed, 6 sites) in Kentucky, and Callahan Creek (15 sites), Roaring Fork (9 sites), and Looney Creek (Pigeon Creek watershed, 8 sites) in Virginia (Fig. 1). Streams were selected to represent four conditions of interest in terms of primary discharge types. Condition 1 (“mining and UHW combination”) is represented by Roaring Fork and Callahan Creek, which both have significant surface mining activities in the headwaters as well as UHW discharges from 20 to 40 homes downstream of mining activities. Condition 2 (“mining only”) is represented by Looney Creek Virginia, which also has surface mining activity in the headwaters, but no known UHW discharges; residences along this creek have municipal sewage service provided by the Town of Appalachia. There is also an historic underground mine discharge near the Looney Creek, Virginia headwaters. Condition 3 (“UHW only”) is represented by Yocum Creek, which has UHW discharges (3–10 homes) but minimal mining activity in the headwaters. Finally, Condition 4 (“reference”) is represented by Looney Creek in Kentucky, which has no current mining activity in the headwaters, and no UHW discharges; there is a municipal wastewater treatment facility discharge point from the towns of Benham and Lynch at the furthest downstream point on the sampled reach of this creek.

2.2. Sample collection

Water quality samples were collected monthly from August 2012 to August 2014 (n = 25) at the longitudinally distributed sites along each stream reach (Fig. 1). Sites were chosen to represent the stream continuum from headwaters to their confluence with higher order reaches, with additional sites selected to highlight inputs from mining discharges and UHW influences. It is worth noting that these study sites were not subject to significant influences from acid mine drainage, as the surface geology of the area is primarily comprised of carbonate formations, lending elevated alkalinity that results in typical stream pH values >7.0.

In-stream measurements of conductivity (µS/cm), temperature (°C), pH, and dissolved oxygen (mg/L) were taken at each site on a monthly basis with an YSI Quatto Pro Plus (YSI, Yellow Springs, OH; USA), with samples for laboratory analyses concurrently collected in sterilized, acid-washed 250 mL polypropylene bottles. The YSI Quattro Pro Plus was tested monthly before use and calibrated as needed (at minimum quarterly) using triple point pH calibration, single point DO calibration in water-saturated air, and single point conductivity calibration using a certified conductivity standard.

Grab samples were transported on ice to the Biological Systems Engineering Water Quality Laboratory in Blacksburg, VA. A 10 mL subsample was acidified with nitric acid (2%) for ICP-MS analysis promptly after field sampling excursions. In cases of high turbidity and particulate matter (i.e., trips following rain events), samples were syringe filtered using a 0.45 µm Whatman PES filter prior to ICP-MS analysis to prevent instrument malfunction. Table 1 includes maximum recoverable limits as well as minimum detection limits for the ICP-MS analysis conducted on these samples. ICP-MS analysis reported S as SO₄²⁻ and Cl as chloride. Data QA/QC included replacing values below minimum detection levels with zero values, as well as checking for anomalies in ICP-MS detection (spikes in observations that indicate problems with the ICP-MS sweep). Because samples were generally unfiltered, results represent total levels of inorganic ions as opposed to dissolved levels. Field blanks were collected and analyzed during each sampling trip to confirm no background sources of contamination.

Alkalinity was determined using a HACH Alkalinity Test Kit (Model AL-DT) on an additional 100 mL subsample. The measured alkalinity could then be used to approximate HCO₃⁻, since alkalinity in the study streams is dominated by carbonates and measured pH values were generally in the neutral to slightly basic range (i.e., pH 7.0–8.3) such that virtually all carbonate is present as HCO₃⁻.

2.3. Statistical analyses

Ward (1963) hierarchical clustering and principal component analysis (PCA) (Legendre and Legendre, 2012b) were used to determine if study sites grouped by pollution source (mining only, UHW only, a combination of the two, or reference conditions) based on measured ion concentrations using all ions shown in Table 1 (Legendre and Legendre, 2012a). No a priori grouping was used in this analysis, thus groups formed based on ion concentrations alone. Appropriate statistical methods were used to address inherent pseudoreplication due to longitudinal sampling. Wards’ hierarchical clustering is an Agglomerative algorithm that begins with all sites as individual groups and iteratively forms larger groups (or clusters) by minimizing the total within-group
variance until all sites become a single group. A dendrogram can be plotted to visualize this bottom-up agglomeration process and tell the user the optimal number of clusters. PCA is an ordination technique used to combine the inorganic ions into principal components that capture the majority of variation in the data. Only the first few principal components with large eigenvalues (empirically >1) are kept, so the dimension of data space can be reduced. A scree plot along with eigenvalue magnitudes were used to determine component inclusion in analysis with a loading cutoff of 0.30 used to consider important ion contributions to each component (Legendre and Legendre, 2012b). Another common practice is to draw a PCA biplot which combines the scores (showing the relations among observations in the 2-dimension space) and a loading plot (showing relations among variables, and relations between original variables and principal components) (Legendre and Legendre, 2012b). Because PCA requires normality in datasets, ICP-MS ions were normalized prior to analysis using a log-normal transformation (Shapiro–Wilk testing indicated the raw data to be non-normal); means of ion concentrations were used for each site (n = 25 observations at each site). For the purpose of preventing multicollinearity in the data, conductivity was excluded from the analysis as it is an inherent measure of the sum of all ions. R statistical programming language v2.15.1 was used for all analysis (R Foundation for Statistical Computing, 2012).

Using the groups observed in CA, PCA biplots were used to color sites based on group, illustrating the location of those groups in data space and allowing interpretation of how the individual ion concentrations were important in terms of group membership. As an evaluation metric, classification and regression tree (CART) analysis was used to determine the misclassification rate of group membership in CA. CART is a recursive tree algorithm process, but can accept data with pre-determined groups and run a tree algorithm to determine how accurately those group labels describe observations by recursively running variables through the tree algorithm, beginning with the data as one large group and splitting the dataset into smaller groups while minimizing variance within groups (Fabricius, 2000).

In concert with CART, the Random Forest algorithm was used to further understand the importance of individual ions in the grouping process. This technique creates numerous tree structures using processes similar to CART; however, some data is held from the analysis and subsequently used to evaluate each tree (out-of-bag validation). The variable importance and model accuracy of each tree is saved, then variable importance is determined over the entire collection of trees (i.e., “forest”); this is generally considered a more robust process than the singular tree development processes used in CART (Breiman, 2001). From this process, further validation is provided for ion importance and thus characterization of how ions drive group membership. The Random Forest method was used to grow 100,000CART-style trees, for which p ≥ 3 variables (where p = 18 is the number of total variables in Table 1) were tested at each split in the tree algorithm to determine those variables that most affect the accuracy of splits that determine group membership. The importance of the different inorganic ions was reviewed over all 100,000 trees (the “forest”) with the response variable being the group classification determined from CA. These processes were completed using the tree package (CART) and the Random Forest package in R (Liaw and Wiener, 2002).

As a final step, permutational MANOVA (PERMANOVA) testing was performed on the CA groups while controlling for stream location. This testing procedure was developed to allow ANOVA to be performed on non-parametric, multivariate data with minimal assumptions and datasets lacking replication; the procedure returns a p-value and F-statistic analogous to Fisher’s F-ratio by using either a dissimilarity matrix or symmetric distance matrix along with a permutation procedure to generate a p-value. The procedure also allows for the inclusion of multiple factors either as additive effects, or as partitioned effects (Anderson, 2001). By treating sites as replicates within a single factor (stream location), the previous concerns of pseudoreplication via longitudinal sampling could be addressed and group differences statistically verified. PERMANOVA testing was completed using the vegan package in R (R Foundation for Statistical Computing, 2012). Last, a correlation matrix was used to review the relationship between

Table 2

<table>
<thead>
<tr>
<th>Parameter</th>
<th>CC (n = 350)</th>
<th>LCY (n = 150)</th>
<th>LCVA (n = 200)</th>
<th>RF (n = 225)</th>
<th>YC (n = 200)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Al</td>
<td>69.51</td>
<td>48.93</td>
<td>53.61</td>
<td>45.61</td>
<td>77.76</td>
</tr>
<tr>
<td>As</td>
<td>0.07</td>
<td>0.03</td>
<td>0.05</td>
<td>0.16</td>
<td>0.06</td>
</tr>
<tr>
<td>Bicarbonate</td>
<td>982</td>
<td>551</td>
<td>1080</td>
<td>884</td>
<td>1833</td>
</tr>
<tr>
<td>Ca</td>
<td>61838</td>
<td>20643</td>
<td>63174</td>
<td>90691</td>
<td>25999</td>
</tr>
<tr>
<td>Cl</td>
<td>3018</td>
<td>7097</td>
<td>6303</td>
<td>1852</td>
<td>5104</td>
</tr>
<tr>
<td>Co</td>
<td>0.23</td>
<td>0.09</td>
<td>0.17</td>
<td>0.52</td>
<td>0.14</td>
</tr>
<tr>
<td>Conduct</td>
<td>6.30</td>
<td>246</td>
<td>692</td>
<td>868</td>
<td>371</td>
</tr>
<tr>
<td>Cr</td>
<td>0.59</td>
<td>0.43</td>
<td>0.7</td>
<td>0.56</td>
<td>0.68</td>
</tr>
<tr>
<td>Cu</td>
<td>0.67</td>
<td>0.19</td>
<td>0.39</td>
<td>0.54</td>
<td>0.50</td>
</tr>
<tr>
<td>DO</td>
<td>9.7</td>
<td>9.7</td>
<td>8.2</td>
<td>9.4</td>
<td>9.6</td>
</tr>
<tr>
<td>Fe</td>
<td>165.12</td>
<td>92.92</td>
<td>169.81</td>
<td>170.4</td>
<td>104.28</td>
</tr>
<tr>
<td>K</td>
<td>4655</td>
<td>2352</td>
<td>4609</td>
<td>5521</td>
<td>2918</td>
</tr>
<tr>
<td>Mg</td>
<td>26737</td>
<td>7817</td>
<td>29269</td>
<td>45954</td>
<td>9734</td>
</tr>
<tr>
<td>Mn</td>
<td>25.13</td>
<td>12.6</td>
<td>912.8</td>
<td>62.32</td>
<td>10.4</td>
</tr>
<tr>
<td>Na</td>
<td>62419</td>
<td>30436</td>
<td>75013</td>
<td>69762</td>
<td>60208</td>
</tr>
<tr>
<td>Ni</td>
<td>1.39</td>
<td>0.36</td>
<td>0.79</td>
<td>1.64</td>
<td>0.44</td>
</tr>
<tr>
<td>P</td>
<td>5.27</td>
<td>4.55</td>
<td>6</td>
<td>2.62</td>
<td>5.32</td>
</tr>
<tr>
<td>pH</td>
<td>8.2</td>
<td>8</td>
<td>8</td>
<td>8.1</td>
<td>8.3</td>
</tr>
<tr>
<td>Se</td>
<td>1.75</td>
<td>3.6</td>
<td>1.1</td>
<td>2.29</td>
<td>2.0</td>
</tr>
<tr>
<td>Si</td>
<td>2379</td>
<td>2060</td>
<td>2753</td>
<td>2335</td>
<td>2407</td>
</tr>
<tr>
<td>SO4</td>
<td>248488</td>
<td>54178</td>
<td>223434</td>
<td>403868</td>
<td>68081</td>
</tr>
<tr>
<td>TEMP</td>
<td>12</td>
<td>11.3</td>
<td>12.8</td>
<td>12.6</td>
<td>11.2</td>
</tr>
<tr>
<td>Zn</td>
<td>6.02</td>
<td>5.99</td>
<td>5.14</td>
<td>6.97</td>
<td>4.96</td>
</tr>
</tbody>
</table>

* All values reported in ppb except conductivity (µS/cm), pH, DO (mg/L), and temperature (°C).

* n-values for bicarbonate are CC (n = 168), LCY (n = 72), LCVA (n = 96), RF (n = 108), YC (n = 98).
field-observed conductivity values and the lab-analyzed ion concentrations in order to review connections of specific ions and the broad measure of conductivity.

3. Results

Mean and standard deviation values for all measured ions and water quality parameters in the five study watersheds are provided in Table 2. A cursory review of this data reveals two general trends: ionic concentrations are lower in the reference watershed (with the exception of chloride) versus watersheds with UHW and/or significant mining activities, and conductivity is highest in watersheds with mining activity (CC, LCA, RF). Additionally, the raw data show that within-watershed values are highly variable; this indicates that water quality is site-specific. Using cluster analysis (CA), sites could be categorized into five groups on the basis of ion concentrations (Fig. 2). Based on knowledge of field conditions, three of these groups correspond very well to the categorical classifications of streams by known primary discharges (i.e., UHW alone, combination of mining and UHW, and reference); and the other two groups correspond to mining-influenced sites, with one group being associated with known surface mine discharges and the other being associated with a known underground discharge. Groups are subsequently referred to as “surface mining”, “underground mining”, “UHW”, “reference”, and “combination,” where “combination” refers to sites downstream from both UHW and surface mining influences.

CART analysis, run to determine classification accuracy from CA, supported these groupings and indicated a misclassification rate of only 2% among the sampling sites. Subsequent Random Forest algorithm results illustrated the inorganic ions important in group formation. The mean decrease in model accuracy related to the permutation of the different variables within the suite of inorganic ions studied is provided in Fig. 3. The most important ions are at the top of the y-axis; permuting them causes a large change in model accuracy. Importance decreases for ions listed near the origin of the plot. Based on this analysis, Ni, Mn, Mg, Ca, K, and Si appear to be the most important ions, with pairs of ions showing increasingly less influence after this group (HCO\textsubscript{3}\textsuperscript{-}, SO\textsubscript{4}\textsuperscript{2-}, then P and Se). A distinct break in the variable importance plot indicates a significant loss of importance.

PCA resulted in the selection of the three primary components (PC1, PC2, and PC3) as in Fig. 4 (Table 3 displays variable loadings and PC eigenvalues). Resultant groups from the CA analysis are differentiated by color; small spheres represent individual sample sites, while a large central sphere for each group represents the group centroid in data space. All component axis vectors point in the positive direction. PC1 explained 37.5% of the variance with significant contributions from Se, Mg, SO\textsubscript{4}\textsuperscript{2-}, K, Ca, and Ni along the negative direction (moving from reference towards surface mining influence). PC2 explained 19.4% of the variance and is associated with significant contributions from Na, Si, P, Cr, and HCO\textsubscript{3}\textsuperscript{-} in the positive direction (moving towards underground mining and UHW influence). The third PC explained an additional 10.8% of the variance with significant contributions from Mn in the positive direction (towards underground mining influence), while Zn, Al, Fe, and Cu increased in the negative direction (towards UHW influence). Table 3 shows the eigenvalues as well as the variable

![Cluster Dendrogram](image)

**Fig. 2.** Grouping dendogram results from cluster analysis.

![Variable importance plot](image)

**Fig. 3.** Variable importance plot for random forest analysis.
loadings for each inorganic ion, with values above the cutoff rate of 0.30 highlighted in bold. Taken with Fig. 4, these values explain how the different inorganic ions push sites into the different groups along PC1, PC2, and PC3. For instance, moving in the negative direction along the PC1 axis, increasing relative concentrations of Ca, Mg, Ni, K, Se, and SO$_4^{2-}$ push sites from the reference group to the combination of (surface) mining and UHW group, and then to surface mining only group.

PERMANOVA testing allowed rejection of the null hypothesis that the groups came from a homogenous distribution (p = 0.001) with an $R^2$ of 0.57 and a pseudo-F statistic of 13.27. Therefore, the grouping detected by cluster analysis (CA) explained 57% of the overall variance. This indicates that while the pseudoreplication introduced by the sampling design may lend effects to the grouping observed in CA, the inorganic ions themselves will explain the groupings most effectively. Further, the pseudo-F statistic indicates that the dispersion of group centroids is much larger than the dispersion within the groups themselves, lending support to the CA results.

The results of correlational analysis between observed conductivity at the sample sites and measured inorganic ion concentrations are shown in a correlation matrix in Fig. 5. The correlation matrix suggests that elevated conductivity in these watersheds is most likely a product of increasing Mg, K, Ca, and SO$_4^{2-}$ compounds, as these ions correlate positively with each other and conductivity. To a lesser extent, Se, Mn, and Ni are also correlated with elevated conductivity and with the aforementioned ions most strongly associated with an increase in conductivity. Some additional ions (i.e., Zn, As, Cr, Fe, Cu) appear to be weakly correlated with conductivity. Quantification of the most important ions shown in the variable importance plot (Fig. 3) is provided by Table 4, which provides a summary of the typical values of each constituent in each watershed. For clarity, the highest concentration of each significant element across groups is shown in bold.

### 4. Discussion

Attempts to describe and predict stream chemistry in the Appalachian region are challenging and complex because of the wide array of inputs, including discharges from various current and legacy surface and underground mining activities (Farag et al., 1998; Merovich et al., 2007; Cravotta, 2008; Lindberg et al., 2011; Timpano, 2011; Agouridis et al., 2012), atmospheric sulfur deposition associated with acid rain (Fitzhugh et al., 1999; Chen and Lin, 2009), and communities with UHW discharges. In cases where mining activity is a driver in altered stream chemistry, the inorganic constituents can vary, and are often driven by the local geological formations found in concert with coal seams. General carbonate formations and coal seams can contain metals such as Ni, Cu, Zn, As and Se (Friedrich et al., 2011; Hartland et al., 2011, 2012; Friedrich and Catalano, 2012; Brown et al., 2014), but relative

<table>
<thead>
<tr>
<th>Parameter</th>
<th>PC1</th>
<th>PC2</th>
<th>PC3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Magnesium (Mg)</td>
<td>0.356</td>
<td>-0.074</td>
<td>0.125</td>
</tr>
<tr>
<td>Manganese (Mn)</td>
<td>-0.178</td>
<td>0.238</td>
<td>0.326</td>
</tr>
<tr>
<td>Nickel (Ni)</td>
<td>-0.311</td>
<td>-0.139</td>
<td>-0.115</td>
</tr>
<tr>
<td>Phosphorus (P)</td>
<td>0.014</td>
<td>0.374</td>
<td>-0.134</td>
</tr>
<tr>
<td>Potassium (K)</td>
<td>-0.355</td>
<td>-0.056</td>
<td>0.008</td>
</tr>
<tr>
<td>Selenium (Se)</td>
<td>-0.300</td>
<td>-0.186</td>
<td>-0.131</td>
</tr>
<tr>
<td>Silicon (Si)</td>
<td>-0.063</td>
<td>0.426</td>
<td>0.155</td>
</tr>
<tr>
<td>Sodium (Na)</td>
<td>-0.167</td>
<td>0.406</td>
<td>0.013</td>
</tr>
<tr>
<td>Sulfur as sulfate (SO$_4$)</td>
<td>-0.364</td>
<td>-0.069</td>
<td>0.057</td>
</tr>
<tr>
<td>Zinc (Zn)</td>
<td>-0.077</td>
<td>-0.063</td>
<td>-0.316</td>
</tr>
</tbody>
</table>
concentrations can vary greatly by location. The presence of chemicals of particular ecological and human concern, such as As and Se, is also geology-dependent as their occurrence with pyrites and other minerals associated with coal seams varies throughout the region (Diehl et al., 2012).

A single relevant study could be found that addressed the effects of sewage on streams in the Appalachian region (Goudreau et al., 1993), though it concerned wastewater treatment plant effluent as opposed to UHW discharges. While some studies have categorized the additive effect of general urbanization in mining-impacted watersheds (Freeman and Schorr, 2004; Clinton and Vose, 2006; Hedrick et al., 2010; Merriam et al., 2011), this study is the first known to include untreated household wastewater (UHW) as a factor in stream chemistry analysis. Because a primary regulatory concern in regards to mining influence is elevated conductivity and metals contamination, inorganic ions are a particularly relevant metric by which to compare stream conditions from these different pollutant types.

From the analysis above, the primary observation is that the source-driven pollutant contributions to inorganic constituent presence did vary by discharge type and could be summarized by the different groupings of ions through a wide array of statistical approaches. CA indicated that sites downstream of untreated household waste discharges are distinguishable from sites where it is absent, or where mining influence displays an additive effect. Sites with known influence from UHW had an average increase in conductivity from under 300 μS/cm to over 500 μS/cm (Table 4). While several important ions increase from reference conditions to UHW conditions as shown in Table 4, the increase in P (only nutrient included in suite of ions) supports the presence of organic compounds. Specifically, the suite of ions along PC2 in Fig. 4 show that sites move from reference conditions to UHW-influenced conditions with increasing Al, Co, Cu, Fe, P, and Zn. Most of these elements are likely associated with surface runoff from developed home sites; however, elemental P is a known nutrient and is most likely associated with household phosphates as well as the organic pollution constituents of UHW discharges.

Along the “mining influence” axis (PC1), Ni and Se are significant in their magnitude, which is consistent with expectations for increased trace metals with increased influence from surface mining operations. AlkMD-associated ions such as Ca, K, Mg, and SO4\(^2^-\) are also relevant. The interesting observation here is that most of the significant ions are cations as opposed to anions, which may be a result of the neutralizing capacity of local geology and soils (cations are leached from geological formations as AMD is neutralized). It should be noted that none of these ions were observed in mean levels that violated surface water quality standards, but their presence is indicative of source-related conductivity elevation.

CART analysis verified the classification of observed sample sites with a very small misclassification rate. Coupled with Random Forest results, the most important grouping of ions in this process was Ni, Mn, Mg, Ca, K, and Si. This result parallels PC1 and PC2 where AlkMD/trace metals defined surface mining influence along PC1 and Si/Mn distinguished underground mining and UHW influence from reference conditions. Other significant ions included in the PCA analysis appear in two pairs following this first group: HCO\(_3^-\)/SO4\(^2^-\) and Se/P, again indicating ions associated with AlkMD, underground mining discharges, and UHW. These pairs show a mixture of important analytes that formed PC2 and PC3. From these analyses it is clear that several different activities can influence conductivity in these watersheds where mining activities are present.

It is likely that changes in water chemistry associated with surface mining and legacy underground mining discharges are primarily associated with elevated conductivity, but the presence of densely-packed homes in legacy “coal camp” clusters influences conductivity as well. In sites on Yocum Creek and Callahan Creek, UHW-impacted sites were also characterized by field-observed macrophyte blooms and conductivity measures consistently above 300 μS/cm, and often as high as 500–700 μS/cm. Downstream measures in conductivity related to surface mining discharges were commonly in this same range, while values in the extensively-mined Roaring Fork watershed were frequently above 1000 μS/cm (as were those downstream from the legacy underground discharge on Looney Creek, Virginia). While most ions characterizing reference conditions were important because of their low concentrations, chloride actually increased in relation to reference sites. Because the reference stream was located in a hallow where one of two local roads over Black Mountain is located, accumulation of road salts may be a possible explanation for this chloride increase. Elevated salinity is often associated with urbanization and the use of road salt (Kausal et al., 2005; Kaye et al., 2006), clouding the matter of reducing conductivity in mining-impacted watersheds further.

Impaired streams in the central Appalachian region are often characterized by elevated measures of conductivity and reduced measures of benthic macroinvertebrate abundance and diversity; however, a lack of data describing the primary constituents of elevated conductivity renders the mechanistic factors behind species degradation difficult to identify, impeding the design of effective restoration plans. The present study used a variety of advanced statistical techniques (CA, PCA, Random Forest, PERMANOVA, and correlation matrices) to support the hypothesis that patterns of inorganic ions are distinct between reference conditions, UHW-impacted sites, and mining-impacted sites in this sub-region of Central Appalachian Coalfields. In efforts to capture the longitudinal effects of local UHW discharges, multiple sites along each stream were sampled (pseudoreplication); to further support the conclusions of this paper, broader, regional studies should be undertaken where the discharges of interest can be replicated across several watersheds. Using longitudinal sites along a stream gradient can cause grouping results and between-group variation to be dominated by the stream unit itself; based on PERMANOVA testing and known field conditions, this effect was not large enough to confound inorganic ion-driven classification of sites.

It is important to further note that the study watersheds were free of AMD influences and tended towards more carbonate-based mining impacts due to local geology, thus ionic suites will likely

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Cond(^a)</th>
<th>Ni</th>
<th>Mn</th>
<th>Ca</th>
<th>K</th>
<th>Si</th>
<th>HCO(_3^-)</th>
<th>SO4(^2^-)</th>
<th>Se</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference (n = 10)</td>
<td>264</td>
<td>0.36</td>
<td>7.63</td>
<td>22672</td>
<td>2484</td>
<td>2174</td>
<td>113897</td>
<td>56687</td>
<td>0.37</td>
<td>2.76</td>
</tr>
<tr>
<td>UHW (n = 6)</td>
<td>562</td>
<td>0.78</td>
<td>22.39</td>
<td>36248</td>
<td>3553</td>
<td>2441</td>
<td>225903</td>
<td>140666</td>
<td>0.85</td>
<td>11.74</td>
</tr>
<tr>
<td>Combination (n = 11)</td>
<td>583</td>
<td>1.14</td>
<td>32.93</td>
<td>58112</td>
<td>4204</td>
<td>2445</td>
<td>153341</td>
<td>230008</td>
<td>1.57</td>
<td>4.15</td>
</tr>
<tr>
<td>Underground (n = 6)</td>
<td>711</td>
<td>0.77</td>
<td>94.01</td>
<td>64074</td>
<td>4052</td>
<td>2818</td>
<td>238641</td>
<td>225563</td>
<td>0.45</td>
<td>6.10</td>
</tr>
<tr>
<td>Surface (n = 12)</td>
<td>838</td>
<td>1.78</td>
<td>51.89</td>
<td>80749</td>
<td>5718</td>
<td>2269</td>
<td>179700</td>
<td>386434</td>
<td>2.30</td>
<td>2.62</td>
</tr>
</tbody>
</table>

\(^a\) Conductivity reported in μS/cm, all other values in ppb.

Table 4: Mean values for significant inorganic ions and conductivity.
vary based on the sub-region of Appalachia. Surface mining disturbances were distinguished primarily by ions that are components of AlkMD (Mg, Ca, K, SO\textsubscript{4}^{2-} specifically), although some of these cations are also elements found in abundance in soils (Mg, Ca, K). Both Ni and Se were observed as trace levels. Mean Se levels were below USEPA surface water quality standards, which is worth noting as recent studies have shown adverse biological effects correlated with Se and its bioaccumulation (Farag et al., 1998; Clements 2004; Lindberg et al., 2011; Arnold et al., 2014). In general, it is likely that inorganic ions associated with UHW sites are the results of surface runoff associated with urbanization as well as organic compounds in-stream, while mining influence varied with surface/underground sourcing. Singly or combined, all of the observed pollutant sources elevated conductivity, but did so with differing suites of inorganic ions that could easily be detected using ICP-MS procedures. A broad approach to the characterization of conductivity sources will improve understanding of the specific stressors of macroinvertebrate and related critical remediation needs, and encourage a more holistic watershedscale–approach to improve water quality and ecological health across the region.

**Author contributions**

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

**Funding sources**

The Appalachian Research Initiative for Environmental Science (ARIES) provided funding for this research. Views contained herein do not necessarily represent the views of ARIES or its funding partners. More information on ARIES can be found at: [http://www.energy.vt.edu/aries/](http://www.energy.vt.edu/aries/).

**Acknowledgments**

Special thanks are given to the Appalachian Research Initiative for Environmental Science (ARIES) for funding this work; the Laboratory for Interdisciplinary Statistical Analysis (LISA) at Virginia Tech for assisting with statistical analysis; Travis Lowe and Tim Compton at Appalachian Technical Services, Inc. in Wise, Virginia for field support; and to several students at Virginia Tech who contributed countless hours of lab, field, and prep work including Vickie Nyström, Sara Gotkulk, RORY COFFEY, and Tyler Faulkner.

**References**


