Watershed zonation through hillslope clustering for tractably quantifying above-and-belowground watershed heterogeneity and functions

Haruko M. Wainwright[1,2], Sebastian Uhlemann[1], Maya Franklin[1], Nicola Falco[1], Nicholas J. Bouskill[1], Michelle E. Newcomer[1], Baptiste Dafflon[1], Erica Woodburn[1], Burke J. Minsley[3], Kenneth H. Williams[1,4], Susan S. Hubbard[1]

[1]Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA
[2]University of California, Berkeley, CA 94720, USA
[3]U.S. Geological Survey, Denver, CO 80225, USA
[4]Rocky Mountain Biological Laboratory, Crested Butte, CO, 81224, USA

Correspondence to: Haruko M. Wainwright (hmwainwright@lbl.gov)

Abstract. In this study, we develop a watershed zonation approach for characterizing watershed organization and functions in a tractable manner by integrating multiple spatial data layers. Recognizing the coupled ecohydrogeological-biogeochemical interactions that occur across bedrock through canopy compartments of a watershed, we hypothesize that (1) suites of above/belowground properties co-varying with each other, (2) hillslopes are representative units for capturing watershed-scale heterogeneity, (3) remote sensing data layers and clustering methods can be used to identify watershed hillslope zones having unique distributions of bedrock-through-canopy properties relative to neighbouring parcels, and (4) property suites associated with the identified zones can be used to understand zone-based functions, such as response to early snowmelt or drought, and solute exports to the river. We demonstrate this concept using unsupervised clustering methods that synthesizes airborne remote sensing data (LiDAR, hyperspectral, and electromagnetic surveys) along with satellite and streamflow data collected in the East River Watershed, Crested Butte, Colorado, USA. Results show that, (1) near-surface bedrock electrical resistivity (top 20 m) and geological structures are significantly correlated with surface topography and plan species distribution, (2) elevation and aspect are independent controls on plant and snow signatures, (3) we can define the scale of hillslopes at which the hillslope-averaged metrics can capture the majority of the overall variability in key properties (such as elevation, net potential annual radiation and peak SWE) and (4) K-means, hierarchical clustering, and Gaussian mixture clustering methods generate similar zonation patterns across the watershed. Using independently collected data, it is shown that the identified zones provide information about zone-based watershed functions, including foresummer drought sensitivity and river nitrogen exports. The approach is expected to be extensible to other sites and generally useful for guiding the selection of hillslope experiment locations and informing model parameterization.

1

Formatted: Bottom: 0.93"

Deleted: approach

Deleted: neighboring

Deleted: associated

Deleted: hillslope-average elevation and slope are significantly correlated with

Deleted: the correlation between hillslope-averaged above- and below- ground properties are significantly higher than pixel-by-pixel correlation

1 Introduction

Predictive understanding of watershed functions is often hindered by the heterogeneous and multiscale fabric of watersheds (e.g., Peters-Lidard et al., 2017). Heterogeneity exists within each of the watershed compartments, including aboveground compartments (i.e., plant species distribution and plant dynamics, topography) and belowground compartments (i.e., soil and bedrock structures/properties). Such watershed *patterns* influence ecohydrological and biogeochemical *processes*, which in turn affect watershed *functions* and create emerging *patterns* as feedback (Sivapalan, 2006). Since watersheds consist of diverse biotic and abiotic compartments, watershed functions include diverse signatures, including hydrological (i.e., partition, storage, and release of water), ecological (e.g., species adaptation, productivity), and geochemical (e.g., nutrient cycling, solute export) signatures (Sivapalan, 2006; Wagener et al., 2007). There is a multiscale nature of heterogeneity such that different processes have different characteristic scales. Watershed hydrology modeling studies typically use the grid size from 100 m to 1 km (e.g., Foster et al., 2020; Maina et al., 2020), whereas soil moisture is known to vary on the order of several or several tens of meters (e.g., Engstrom et al., 2005; Wainwright et al., 2015), and one meter or less for biogeochemical dynamics (e.g., Burt and Pinay, 2005; Groffman et al., 2009).

There have been extensive studies to investigate how the heterogeneous watershed organization influences water, energy, and nutrient cycling and their fluxes (e.g., Peters-Lidard et al., 2017). There are two directions for tackling this problem: a bottom-up Newtonian approach or a top-down Darwinian approach. The Newtonian approach is a reductionist approach to describe a system by a set of mass/energy/momentum conservation equations with spatially variable parameters. Recently, integrated hydrological and reactive transport models have been successfully implemented to describe and predict watershed behaviors from hillslope to watershed scales (e.g., Maxwell and Kollet, 2008; Li et al., 2017). In addition, the Hydrological Response Unit concept has been used to classify the landscapes based on spatial datasets (e.g., landcover types, elevation, and soil maps) and to parameterize hydrological models (Flugel, 1997; Aytaç, 2020). On the other hand, the Darwinian approach identifies rules or organizing principles governing spatial patterns of complex datasets and defines watersheds as self-organized and co-evolved units by watershed functional traits (McDonnell et al., 2007). Catchment scaling and similarity concepts have been used to synthesize the catchment datasets across scales and to classify catchments (Wagener et al., 2007; Thompson et al., 2011; Sawicz et al., 2011; Krause et al. 2014).

In parallel, there have been significant advances recently in understanding and quantifying the watershed-scale heterogeneity of bedrock-to-canopy terrestrial compartments, which regulate water and nutrient cycling and their exports; particularly through the critical zone observatory (CZO) network (Brantley et al., 2017). In particular, Pelletier et al. (2018) have highlighted the control of slope aspects on ecosystem and critical-zone systems, finding that, for example, in water-limited systems, the north-facing slopes have less evapotranspiration and hence higher soil moisture, deeper weathering, and larger nutrient retention in soil (e.g., Hinckley et al., 2014). Such advances are largely attributed to a variety of spatially extensive

Deleted: Studies have suggested that the relevant spatial scale for hydrological heterogeneity governing fluxes is on the order of 100 m

Deleted: Wood et al., 2011

Deleted: for

Deleted: it is

Deleted: and continental

Deleted: datasets, and

characterization technologies across bedrock-to-canopy compartments, which provide various *patterns*. High-resolution DEM
80 from LiDAR have been applied to better understand the relationship between geomorphology and hydrology (Prancevic and
Kirchner, 2019), as well as to measure snow depths and snow-water-equivalent (SWE) over a basin scale (Painter et al., 2016).
LiDAR data was also able to inform near-surface soil properties (Patton et al., 2018; Gillin et al., 2015), hydrological
connectivity (Jensco et al., 2009) and biogeochemical hotspots (Duncan et al., 2013). In addition, hyperspectral remote sensing
can map plant traits (e.g., Asner et al., 2015), leaf water content (e.g., Colombo et al., 2008), leaf chemistry (e.g., Feilhauer, et
85 al., 2015) and other properties, which are also proxies for soil biogeochemistry (Madritch et al., 2014). At the same time,
geophysics has been extensively used to characterize the subsurface structure and to estimate soil and bedrock properties.
Surface geophysics has been used to measure bedrock depth, weathering zone thickness, and other properties (e.g., de Pasquale
et al., 2019), contributing to hillslope-scale hydrological characterization and modeling. Surface ERT and seismic data have
also revealed the influence of tectonic stresses and hydrological processes on bedrock fracturing and weathering (Rempe and
90 Dietrich, 2014; St. Clair et al., 2015). Airborne geophysics—particularly airborne electromagnetic (EM) surveys—was
originally developed for mineral exploration, but is now increasingly used for water-resources applications (e.g., Barfod et al.,
2018; Ball et al., 2020).

Despite these advancements, there are still challenges in associating watershed *functions* to heterogenous watershed *patterns*.
95 It remains a challenge to integrate multitype and multiscale datasets, including ground-based point measurements and airborne
or satellite remote-sensing datasets. Although the aboveground properties, such as topography and plant characteristics, can
be mapped over the watershed scale, the subsurface variability is still difficult to map over that scale, which is one of the
biggest uncertainties in hydrology (Fan et al., 2019). In addition, even though hillslope-scale characterization and experiments
(such as tracer tests) can be extremely useful for providing detailed information about watershed functions (e.g., Hinkley et al,
100 2014); it is difficult to select several hillslopes for such intensive characterization, or to gauge the representativeness of one
hillslope for an entire watershed.

In this study, we develop and test the concept of a zonation approach for tractably characterizing the organization of a
watershed based on multiple spatial data layers, and how these characteristic *patterns* aggregate to predict watershed *functions*.
105 The clustering-based zonation approaches have emerged recently as effective spatial data integration methods that use spatial
clustering to identify regions or zones that have unique distributions of heterogeneous properties and key functions relative to
neighboring regions (Hubbard et al., 2013; Wainwright et al., 2015; Hermes et al., 2020). For watershed zonation, we consider
a hillslope as a fundamental unit for watershed hydrology and element cycling, funneling water and elements from the ridge
to the river (Fan et al., 2019), and also representing aspect controls on critical zones (Pelletier et al., 2018). We follow Band
110 (1989) and Band et al. (1991; 1993) to investigate the appropriate scales of hillslopes for capturing the watershed heterogeneity,
while limiting the internal variance within hillslopes.

We then hypothesize that (1) a suite of above/belowground compartments—plants, topography, soil, bedrock—are co-varying with each other through complex ecohydrological-biogeochemical interactions, (2) hillslopes are representative units for capturing watershed-scale heterogeneity, (3) we can identify a group of hillslopes or watershed-scale zones that have unique distributions of bedrock-through-canopy properties relative to neighboring parcels, and (4) the identified zones can capture the variability of key watershed functions. We demonstrate our approach using the airborne and spatial datasets collected in the East River watershed region near Crested Butte, Colorado, USA (Hubbard et al., 2018). We apply and compare multiple clustering methods to understand the characteristics, commonality, and differences among each method. Finally, we validate the zonation hypothesis based on the datasets that define key watershed functions, including drought sensitivity of plant productivity and water/nitrogen export.

2 Site and Data

We consider the domain of approximately 15 km by 15 km (Figure 1) near Gothic, Colorado, USA, which is the same area used in a recent study (Wainwright et al., 2020). As described in Hubbard et al. (2018), the domain includes four catchments, including the East River, Washington Gulch, Slate River and Coal Creek. It is a part of the Elk Range in the Rocky Mountains, with elevation from 2800 m to 4000 m (Figure 1a). The major land-cover types (NLCD 2011) are rock outcrop (12%), evergreen forest (29%), deciduous forest (18%), grassland (30%), and woody wetland (6%). Geology within the domain is diverse, including Paleozoic, Mesozoic, and Cenozoic sedimentary rocks (silt- and sandstones of the late Permian Maroon Formation; shales and sandstones of the upper Cretaceous Mancos Shale and Mesaverde group, respectively; silt- and sandstones of the Eocene Wasatch Formation) and Miocene igneous intrusive rocks of predominantly granodioritic composition (Gaskill et al., 1991).

The spatial data layers include the USGS landcover map (NLCD 2011), the digital geological map of Colorado (Green et al., 1992), and the soil texture maps (%clay and %sand) from the POLARIS database (Chaney et al., 2016; 2019). In addition, we used four airborne datasets (Texts S1): an airborne electromagnetic (AEM) survey acquired in fall 2017 (Minsley and Ball, 2018; Uhlemann et al., submitted; Zamudio et al., 2020), LiDAR and hyperspectral data collected by the National Ecological Observation Network (NEON) team in June 2018 (Chadwick et al., 2020), and NASA Airborne Snow Observatory (ASO) data collected in April 2018 (doi.org/10.5067/M4TUH28NHL4Z).

To test the zonation hypothesis, we used datasets representing two key functions: foresummer drought sensitivity of plant productivity and river nitrogen export. The foresummer drought sensitivity map based on Landsat Normalized Difference Vegetation Index (NDVI) was developed by Wainwright et al. (2020) to represent the plant productivity responses to early snowmelt and subsequent drought conditions in the primary growing season. To create this map, Wainwright et al. (2020) performed the linear regression of the historical peak Landsat NDVI as a function of the Palmer Drought Severity Index in

4

June. They then defined the slope of the linear fit as the foresummer drought sensitivity, which represented the magnitude of peak plant productivity changes with respect to the drought condition in the growing season. Since the satellite images were not used in clustering, they were considered independent datasets. In addition, the annual discharge and nitrogen export were computed from the streamflow and chemistry data at the sub-catchments (Figure 1) within the domain used in Carroll et al. (2018). The nitrogen export was computed in the same manner as Newcomer et al. (2021).
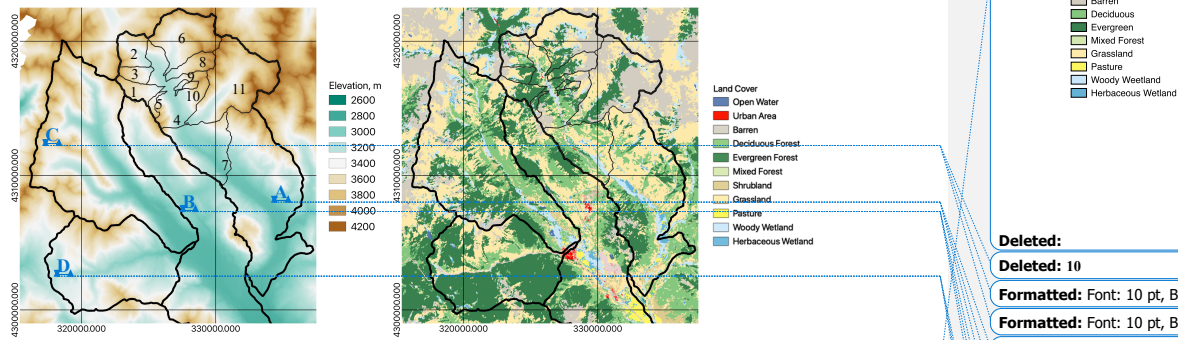
Figure 1. Study domain with (a) elevation and (b) land cover map (USGS NLCD 2011). The black lines are the boundaries of the four watersheds: (A) East River, (B) Washington Gulch, (C) Slate River and (D) Coal Creek. The thin black boundaries are sub-catchments of the East River watershed from Carroll et al. (2018): (1) Rock, (2) East Above Quigley (EAQ), (3) Quigley, (4) East Below Copper (EBC), (5) Gothic, (6) Rustlers, (7) Pumphouse, (8) Bradley, (9) Marmot, (10) Avery, and (11) Copper.

3 Methodology

3.1 Watershed and Hillslope Characteristics

We followed Wainwright et al. (2020) to compute pixel-by-pixel topographic metrics based on the LiDAR Digital Elevation Model (DEM), including slope, Topographic Wetness Index (TWI), and annual net potential solar radiation. We selected metrics based on the previous studies that reported their relevance to soil moisture, soil thickness, water quality and others (Mohanty et al., 2000; Gillin et al., 2015; Lintern et al., 2018). The annual net potential solar radiation – a function of the aspect, slope, and solar angle – is considered as a better metric to represent the intensity of solar radiation than the aspect itself which is circular (0 and 360 degrees are the same; Wainwright et al., 2020). In parallel, we delineated each hillslope based on the DEM, following Noel et al. (2014) and using Topotoolbox (Schwanghart, 2014). The hillslope delineation algorithm first

5

190 identifies the stream segments that are a collection of pixels that have larger flow accumulation (i.e., a larger drainage area) than a certain threshold value, using the flow routing algorithms in Topotoolbox. It then finds the pixels in both sides of each stream segment, and also the drainage area leading to each of these pixels. This process yields two lateral hillslopes in both sides of each stream segment. For the first-order stream, the algorithm also defines the headwater hillslope, which is the drainage area leading to the pixel at the origin of the stream. Although these hillslopes are based on surficial water routing, we

195 assume that the DEM captures near-surface hydrological connectivity as documented by Jensco et al. (2009).

We then defined the characteristics of each hillslope, based on the spatial data layers (Figure S1). We computed the average values of the DEM topographic metrics, AEM-based bedrock resistivity, NEON products (Normalized Difference Vegetation Index, NDVI; Normalized Difference Water Index, NDWI, and biomass) at the peak growing season (2018), peak SWE (2018)

200 and soil texture in each hillslope. In addition, we computed the relief of each hillslope, which was the difference between the minimum and maximum elevation. For the categorical variables such as land-cover types and geology, we computed the percent coverage of each plant type and surface geology in each hillslope. These 17 hillslope features are defined for each hillslope (Table S1).

205 As Band (1989) and Band et al (1991) noted, the hillslope delineation can create different sets of hillslopes, depending on the threshold drainage areas to define the stream segments. For the hillslope metrics defined by the average (e.g., elevation, slope, and others), we evaluated how these averaged metrics can represent the overall variability of the watershed properties, and how this representation changes depending on the different threshold drainage areas. We consider that the variance of pixel-by-pixel properties represents the overall variability over this domain, while the variance of hillslope-averaged metrics

210 represents the variability across the hillslopes. We computed the ratio between the across-hillslope variance and the overall variance, representing how much watershed-scale variability the hillslopes can capture. In addition, as a contrast, we computed the variance of each property within the upscaled pixels, by taking the averaged values in larger pixels compared to the original 9-meter pixel. We computed the ratio between this across-upscaled-pixel variance and the overall variance. In this way, we can investigate the difference between hillslope-based spatial aggregation compared to standard pixel-based upscaling for each

215 of the key watershed properties.

3.2 Cluster-Based Approach to identify Watershed Zones

Based on the hillslope features, we first evaluated the correlations among multivariate above/belowground properties. Although such multivariate co-variability has been analyzed using principal component analysis (Devadoss et al., 2020), we

220 used scatter plots and correlation coefficients in this study because of nonlinearity. We then applied three commonly used unsupervised clustering methods: K-means (KM), hierarchical clustering (HC), and Gaussian mixture models (GMM; Hastie et al., 2001; Kassambara, 2017). We scaled each feature by the mean and standard deviation, and defined the dissimilarity

6

Deleted: We set the threshold of flow accumulation to match the stream observations at major tributaries of the East River (Carroll et al., 2018). …

Deleted: the hillslope

Deleted: is

Deleted: and

Deleted: 5

Deleted: parameters in

between two data points based on the Euclidean distance. The characteristics of each method are described in Texts S2. Multiple methods are often evaluated based on true classes or labeled datasets (Rodriguez et al., 2019), which are not available here. We used a silhouette score, which represented how similar a given data point was to its own cluster compared to other clusters. For GMM, we used the Bayesian Information Criterion (BIC) for selecting the number of clusters.

After the clusters are defined, we transfer the clusters—the group of hillslopes that have similar features—to the spatial map as zones. We identify the common zones across the three methods as well as the zones that differ. We then evaluate the distribution of hillslope features and functions in each zone using box plots to define the characteristics of each zone. The foresummer drought sensitivity (Wainwright et al., 2020) is a watershed function available throughout the domain that allows us to quantify the hillslope-average values in the same manner as other spatial data layers. In addition, we computed the spatial coverage of each zone in each sub-catchment, and compared them to the ratio between the annual nitrogen export and total discharge, which is considered as a key metric of how watersheds retain and lose nutrients (Newcomer et al., 2021).

4. Results

4.1 Hillslope Scales

The ratio between the across-hillslope variance and the overall variance (Figure 2a) is generally high for the elevation, net annual potential solar radiation (radiation), peak SWE and bedrock resistivity up to 0.75, which means that the hillslope averaged metrics capture the watershed-scale variability of these variables and that the within-hillslope variability is small compared to the across-hillslope variability. TWI and NDVI, on the other hand, have a low ratio, which means that the within-hillslope variability is significant. The ratio increases as the drainage area decreases, since the smaller the hillslopes are, the better they capture the small-scale variability. However, the variance ratio of the elevation and radiation reaches a plateau with the drainage area around $10^6 m^2$. This means that the internal variability within hillslopes is limited up to this threshold drainage area, and that the hillslope metrics are representative of the overall variability up to this threshold. Based on this result, we selected 810,000 m^2 as the threshold drainage area in the subsequent analysis.

We can compare this hillslope-based averaging (Figure 2a) with pixel-based averaging/upscaling (Figure 2b). Similar to hillslope-based averaging, the ratio between the across-upscaled-pixel variance and the overall variance increases as the pixel size decreases. The magnitude is similar such that the elevation, peak SWE and bedrock resistivity have a higher ratio, meaning that the upscaled properties can capture the overall variability of these properties. The exception is the radiation, since the pixel-averaged radiation captures only up to 60% of the overall variance, while the hillslope-averaged radiation captures up to 70%. In addition, different from the hillslope average, the ratio for radiation in Figure 2b keeps increasing without reaching a plateau. This means that a representative size or scale does not exist when we use pixel-based upscaling.
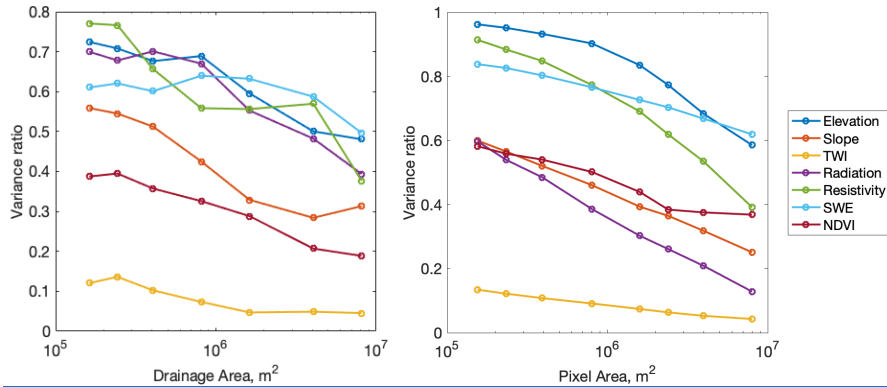
265

Figure 2. Variance ratio (a) between the across-hillslope variance and the overall variance as a function of the threshold drainage area, and (b) between the across-upscaled-pixel variance and the overall variance as a function of the pixel area.

4.2 Watershed Zonation

270 Figure 3 shows that the above- and belowground hillslope features are significantly correlated with each other. In particular, elevation is correlated with all other features except for net annual potential radiation. The hillslopes at higher elevation have steeper slopes and lower TWI, higher bedrock resistivity and higher peak SWE. There are some nonlinearities: TWI increases beyond the linear function at lower elevations. The relationship between peak NDVI and elevation is quadratic, having peaks in mid-elevation (corresponding to around 3300 m). The net annual potential radiation is weakly correlated with slope, TWI,

275 and peak SWE. The correlation coefficients among the hillslope-averaged features are significantly higher than the ones of pixel-by-pixel ones (Figure S2).
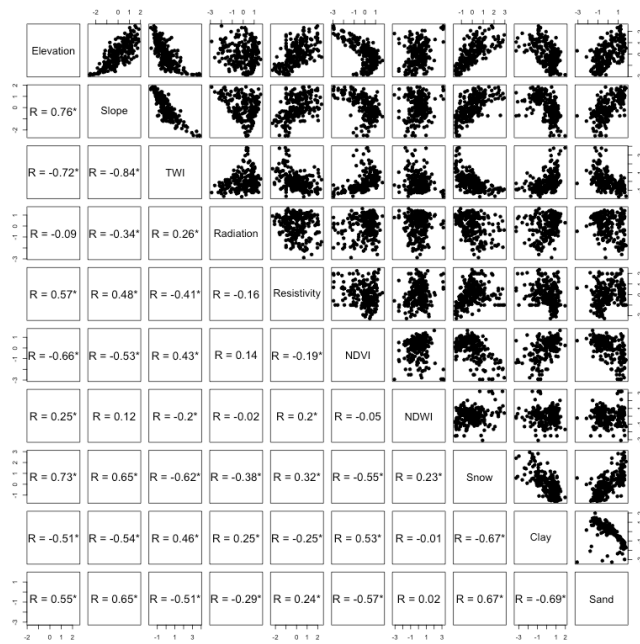
280 **Figure 3. Correlation and scatter plots (Pearson's correlation coefficients) among selected hillslope features (Table S1). The * sign represents p-values < 0.01.**

The clustering results are then mapped spatially as the watershed zones (Figure 4). The number of clusters is six, since the highest silhouette score is high at six clusters across the three methods (Figure S3a), and BIC for GMM is also the highest at 285 six clusters (above four; Figure S3b). To compare the results from these three methods, we first identified six common zones across the three methods that have the overlapping coverage, starting from the GMM-based map as a basis. Zone 7 appears only in the HC result (Figure 4c) and is designated as a separate zone; the hillslopes in Zone 7 are included either in Zone 3 and 4 in the GMM and KM results. The frequency maps (Figure S4) show that most hillslopes are consistently categorized in the same zones among three methods. The zonation maps from GMM (Figure 4a) and KM (Figure 4b) are quite similar, 290 although small fractions of the hillslopes have different designation between the two maps. The zonation map from HC (Figure 4c) is also similar to the ones from GMM and KM except for Zone 7 which is a part of Zone 3 or 4 in the GMM and KM result.
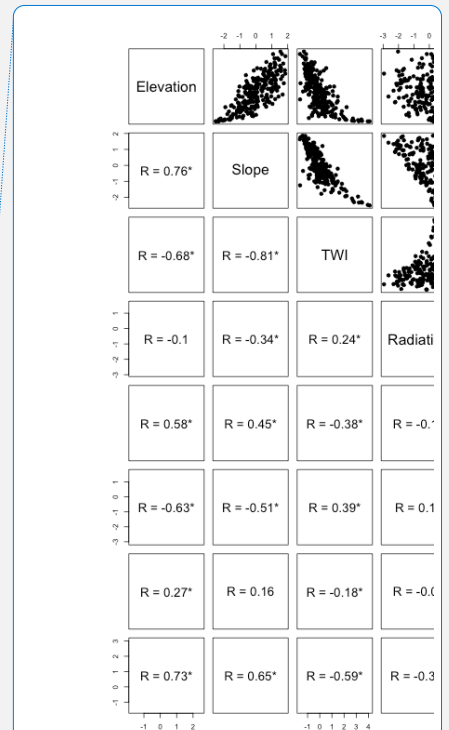
Deleted:

Deleted: 2

Deleted: 3

Deleted: seven

Deleted: GMM showed

Deleted: at seven clusters

Deleted: seven

Deleted: five

Moved (insertion) [4]

Deleted: some

Deleted: ¶
The zones are similar among the three methods, although there are minor differences. Zones 1–6 are common zones across the three methods. …

Deleted: Zone 7 appears in GMM (Figure 3a), but is categorized as Zone 1 by the other methods. Zone 8 appears only in KM (Figure 3b); Zone 8 is mostly a part of Zone 3 in the other methods, a… [1]

Moved up [4]: The frequency maps (Figure S4) show that some hillslopes are consistently categorized in the same zones among three methods. ¶
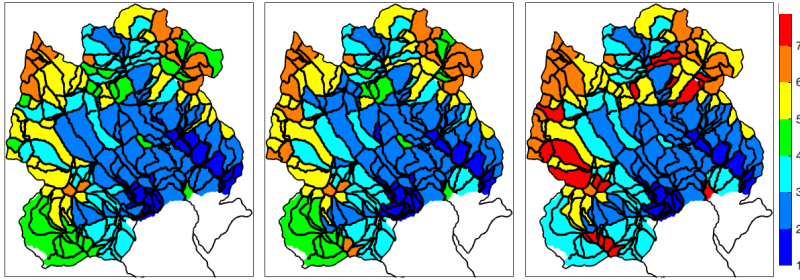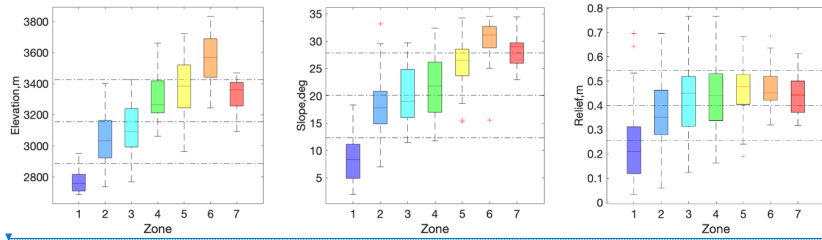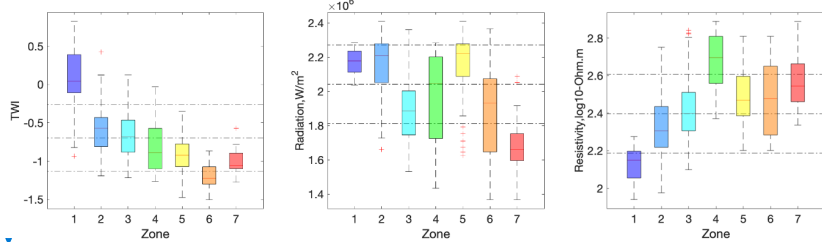
(a)   (b)   (c)

Figure 4. Hillslope zonation based on (a) Gaussian mixture model (GMM), (b) K-means (KM), and (c) hierarchical clustering (HC). Different zones are represented by the cluster index from 1 to 7. There are 6 common zones among the three methods, and 1 additional zone (Zone 7) appears in the HC result (c). The black lines are the hillslope boundaries.

Box plots shows the characteristics of each identified zone (Figures 5 and 6). Although each map has six zones, we use seven zones that are found across the three clustering methods. The statistics are computed by aggregating the zonation from the three methods (Figure S4) such that the zonation of each method has the weight of 1/3. Elevation (Figure 5a) increases from Zone 1 to 6, while Zone 7 is similar to Zone 4. Slope (Figure 5b) shows a similar progression to elevation, which is consistent with the correlation plot (Figure 3), although Zone 7 has steeper hillslopes than expected for its elevation range. Relief (Figure 5c) also has a similar order to the elevation, except that Zones 4 to 7 are similar, because the higher elevation hillslopes are smaller with lower relief, even though the slope is high. TWI (Figure 5d) has an opposite trend to elevation, consistent with the correlation plot (Figure 3). In terms of the net annual potential radiation (Figure 5e), the grassland-dominated hillslopes (Zones 1, 2 and 5) tend to be higher, while the conifer-dominated hillslopes (Zones 3, 4 and 7) are lower except for Zone 4 being around the average. The bedrock resistivity (Figure 5f) is higher at the higher elevation zones (Zones 4 – 7) or an intrusive-rock-dominated zone (Zones 4). NDVI (Figure 5g) is higher in the forest-dominated zones (Zone 2, 3, 4, and 7). Zone 6 has significantly lower NDVI, since it is predominantly in the barren region. NDWI (Figure 5h) is lower in the low elevation zones (Zones 1). Peak SWE (Figure 5i) is associated with the elevation such that it is the highest in Zone 6, which has the highest elevation, and the lower net potential radiation.
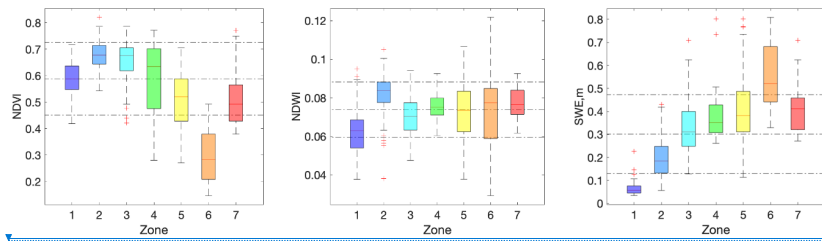
(a) (b) (c)

385

(d) (e) (f)

(g) (h) (i)

Figure 5. Boxplots of hillslope pattern features: (a) elevation, (b) slope, (c) relief, (d) TWI, (e) annual net potential radiation, (f) bedrock electrical resistivity (top 20 m), (g) peak NDVI in 2018, (h) NDWI (at peak NDVI) in 2018, and (i) peak SWE in 2018. The
390 **horizontal lines are the overall mean and plus/minus one standard deviation.**

With respect to geology (Figure 6a-c), the low-elevation zones (Zone 1 and 2) are predominantly shale, while the mid-to-high elevation zones (Zone 3 and 5) have both shale and sandstone. Zone 4, 6 and 7 have a high percentage of intrusive rock coverage. Among the conifer-dominated areas (Zones 3, 4, and 7), Zone 4 is associated with intrusive rock, while Zone 3 is

11
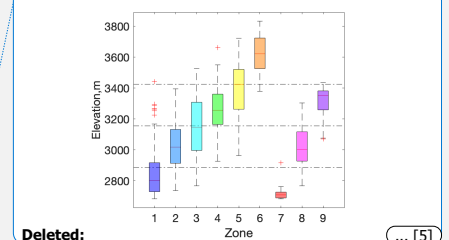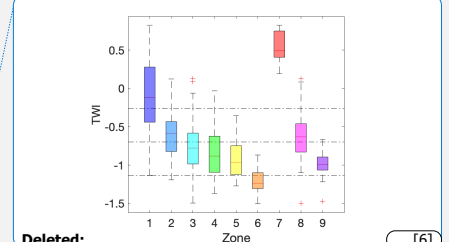
Deleted: [... 5]

Deleted: [... 6]
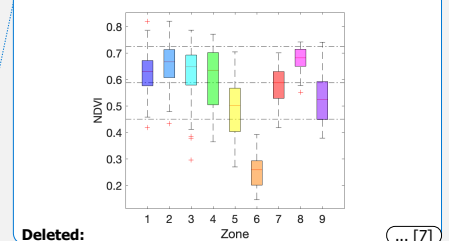
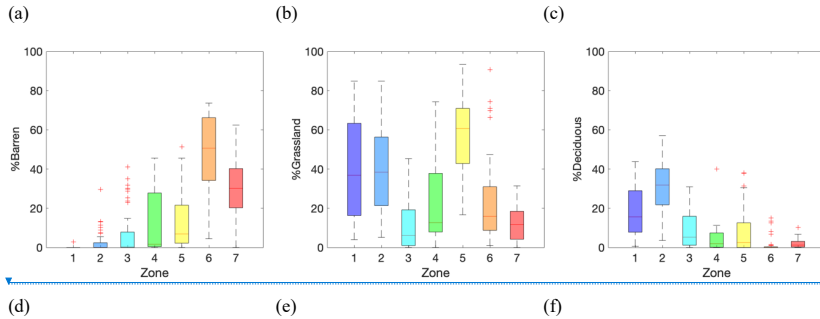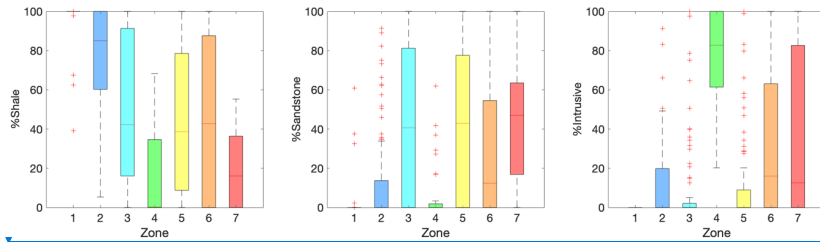**Deleted:** ¶

Deleted: [... 7]
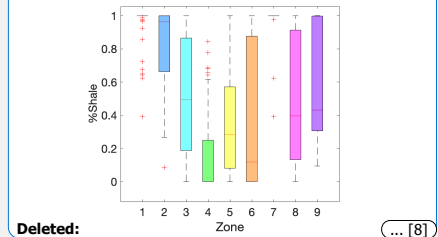
Deleted: 4

Deleted: 5a

Deleted: ; shale coverage (Figure 5a) decreases in the higher-elevation zones that see a concomitant increase in intrusive rocks and hydrothermally altered sandstones (Zones 4, 5, 6, 8), whereas non-altered sandstone regions (Figure 5b) are dominant in mid-elevation hillslopes (Zone 3 and 5).
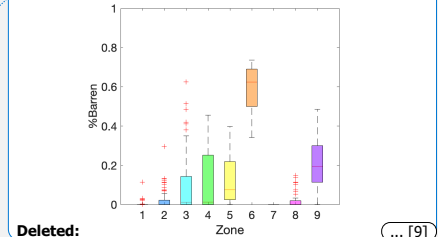
Deleted: 8, 9

associated with shale and sandstone, and Zone 7 have the mixture of these three bedrock types. In terms of land-cover types (Figure 6d-g), the barren area coverage (Figure 6d) increases with elevation from Zone 1 to 6. The grassland region (Figure 6e) exists across all zones, but it is particularly high in the low-elevation zones (Zones 1 and 2) and the higher-elevation south-facing zone (Zone 5). The coverage of deciduous forests (Figure 6f) is higher at lower elevation (i.e., Zones 1 and 2); Zone 2, in particular, has high aspen coverage. The evergreen forest regions (Figure 6g) appear in Zones 3, 4, and 7, which are associated with different bedrock types and slopes as mentioned above. Finally, the wetland (woody and herbaceous combined) coverage is higher in the low-elevation Zone 1. The characteristic of each zone is summarized in Table 1.

420

425

(a)   (b)   (c)

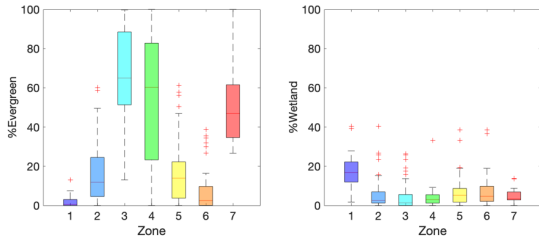(d)   (e)   (f)
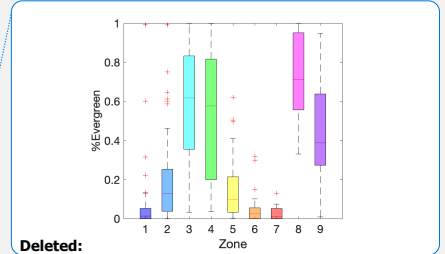
12

(g)   (h)

Figure 6. Boxplots of hillslope pattern features with respect to geology and land-cover types: the percent coverage of (a) shale, (b) sandstone, (c) igneous (intrusive) rock, (d) barren, (e) grassland, (f) deciduous forest, (g) evergreen forest and (h) riparian zone.

460

Table 1. Summary of the features in each zone based on Figure 5 and 6.

Zone	Elevation	Slope	TWI	Resistivity	Radiation	Peak NDVI	Peak SWE	Plant types	Geology
1	Low	Low	High	Low	Mid-High	Mid	Low	Grassland	Shale
2	Low-Mid	Low-Mid	Mid-High	Low-Mid	Mid-High	Mid-High	Low-Mid	Grassland-Deciduous	Shale
3	Mid	Mid	Mid-High	Mid	Low	Mid-High	Mid	Evergreen	Shale-Sandstone
4	Mid	Mid-High	Low-Mid	High	Mid	Mid-High	Mid-High	Evergreen	Intrusive
5	Mid-High	Mid-High	Low-Mid	Mid-High	Mid-High	Low-Mid	Mid-High	Grassland	Shale-Sandstone
6	High	High	Low	High	Mid-Low	Low	High	Barren	Shale-Intrusive
7	Mid-High	High	Low-Mid	Mid-High	Low	Low-Mid	Mid-High	Evergreen	Mixed

4.3 Watershed Functions

Foresummer drought sensitivity (Figure 7a) has an opposite trend to elevation (Figure 5a), with higher sensitivity in lower-

465 elevation zones (Zone 1 and 2), and the lowest in Zone 6. Zone 5 is an exception such that this zone – high-elevation south-facing grassland-dominated hillslopes – has higher sensitivity than the zones that are in a similar elevation range. The conifer-dominated zones (Zones 3, 4 and 7) have lower-than-average sensitivity. Tukey's pairwise comparison metrics suggest that the differences are significant except in the case of conifer-dominated Zones (3, 4 and 7).

470 We found that the magnitude of annual nitrogen exports scaled log-linearly with discharge, albeit with bifurcation among some of the smaller catchments (Figure 7c). The Rock and Gothic subcatchments, which are predominantly within conifer-dominated Zone 3, 4 and 7 (Figure 7b), have a lower nitrogen expected from this simple scaling relationship. By contrast, Zone 2 dominated subcatchments (Marmot, Avery) have a larger nitrogen export than would be expected from this relationship.

13

In addition, we evaluated the association of the N mass export and % coverage of each zone quantitatively. Since the number of points is not large, we divided the N export data into the two groups of the sub-catchments that have the coverage of each zone larger than 50% or less (Figure 7d and e). The association is statistically significant for Zone 2 and for the conifer-dominated zones combined (Zone 3, 4 and 7) with the p-value of less than 0.01. The larger coverage of Zone 2 is associated with a higher N export, while the conifer-dominated zones are associated with a lower export. The individual data points are also shown in Figure S5.
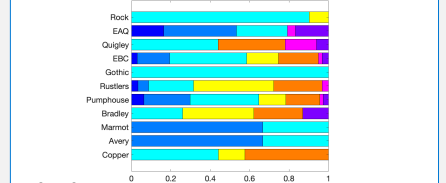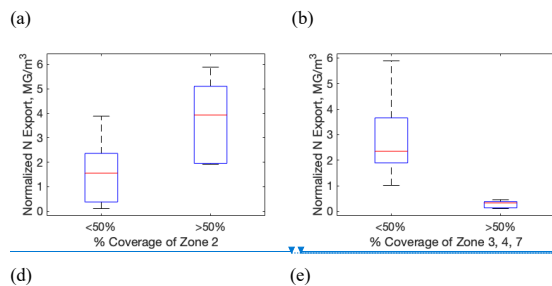

500 (a)   (b)   (c)

(d)   (e)
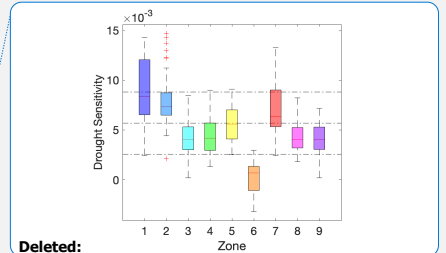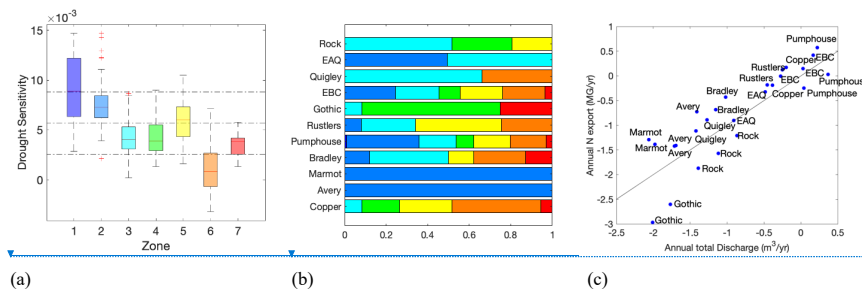
Figure 7. (a) Boxplot of the foresummer drought sensitivity as a function of zones, (b) zonation coverage of each subcatchment, (c) annual discharge vs N export in 2015, 2016 and 2017, and (d, e) normalized N export (N export divided by annual discharge) as a function of the percent coverage of (d) Zone 2 and (e) Zone 3, 4 and 7 as the conifer dominated zone, respectively. In (b), the color code is the same as Figure 3 – 5.

5 Discussion

We first identified the co-variability between topographic metrics, elevation, bedrock resistivity, and plant signatures at the watershed scale. In particular, the near-surface bedrock electrical resistivity is significantly correlated to slope and elevation. The detailed analysis of AEM by Uhlemann et al. (submitted to Scientific Advances) has found that, (1) bedrock resistivity is

14

550 primarily controlled by bedrock type, and (2) within the shale and sandstone, the bedrock resistivity is affected by the extent of hydrothermal alteration. The relationship between topography, elevation, and bedrock strength has been documented extensively (e.g., Selby, 1987; Clarke and Burbank, 2010). Montgomery (2001) showed that, in low-precipitation and low soil production environments, such as the East River, the erosion rate is dependent on rock strength and slope steepness, whereby erosion-resistant rock (i.e., hydrothermally altered sandstones and shales and/or igneous rocks) are on steeper slopes than more

555 erodible rocks. Erosion-resistant rocks also remain at higher elevations, while erodible rock erodes away. This work provides information on how surface topographic features are linked to bedrock variability and could be potentially used to inform bedrock variability, which is considered as one of the largest uncertainties in watershed science (Fan et al., 2019).

In addition, our results show that plant signatures and peak SWE are primarily correlated with elevation. The quadratic

560 dependency of NDVI and aboveground biomass on elevation is consistent with previous studies (e.g., Wainwright et al., 2020); productivity and biomass increases with elevation up to 3300-3500 m because plants have more access to snow-derived water at higher elevation. Above 3500 m, plant productivity is limited by temperature. Net annual potential radiation is not correlated with elevation, but associated with conifer coverage and peak SWE. The conifer forest and high vegetative biomass are associated with north-facing slopes due to higher water availability, which is consistent with previous syntheses studies (e.g.,

565 Pelletier et al. 2018). The spatial variability of peak SWE is also consistent with previous studies being associated with elevation and radiation (Anderson et al., 2016). In addition, the conifer-dominated zones are associated with high-resistivity bedrock, while aspen forests are located within the less resistive bedrock. This is consistent with past studies that conifer trees grow in the shallow bedrock regions, while aspen require the establishment of deeper root systems (e.g., Burke and Kasahara, 2011). Although there have been studies that have reported the co-variability of plants, topographic metrics, and aspect-

570 controlled bedrock weathering (e.g., Pelletier et al, 2018), our analysis suggested an additional control of geological variability.

To capture such co-variabilities, we hypothesized that hillslopes would be appropriate spatial units to aggregate key properties and to classify into zones. However, the hillslopes could be defined at different scales based on the threshold drainage area, as documented by Band (1989) and Band et al. (1991). We computed the variance of hillslope-averaged metrics compared to the

575 overall variance as a measure for quantifying how well hillslopes could capture the overall heterogeneity. We then investigated the effect of the threshold drainage area on this variance ratio. Our results show that, although this ratio decreases as the drainage area increases, it stays flat up to a certain threshold drainage area, particularly for elevation and annual net potential radiation. This suggests that there is a representative scale of hillslopes, up to which the within-hillslope variability is limited, and the hillslope-averaged metrics can capture most of the overall variability across the domain. This representative scale is

580 important, since Pelletier et al. (2019) showed a significant impact of hillslope aspects and slopes on evapotranspiration, weathering, nutrient cycling, and others.

15

Deleted: hydrothermally-altered

Deleted: Our

On the other hand, standard pixel-based upscaling (i.e., taking an average in larger pixels) does not show such a plateau, since larger pixels include a mixture of hillslopes with different aspects. This is consistent with Band et al. (1991) which documented that the hillslope-based partition had much greater efficiency for upscaling than pixel-based scaling by minimizing the within-unit variability and maximizing the ability to capture the watershed-scale variability. Although pixel-based (or grid-based) spatial scaling is still the mainstream for remote sensing and hydrological modeling (e.g., Peng et al., 2017; Foster et al., 2020), our results suggest that the hillslope-based partitioning would be a more effective template for scaling, particularly for mountainous hydrology where aspects and solar radiation play a significant role (Pelletier et al., 2018; Fan et al., 2019).

After the appropriate hillslope scale is defined, we computed the hillslope features (including the hillslope-average of key properties as well as the percent coverage of categorical variables), and applied unsupervised clustering to these features. This clustering process takes advantage of the co-variability between aboveground and belowground properties and reduce the multidimensional parameter space into one-dimensional classes (Hastie et al., 2013). In the East River watershed, the zones are primarily dictated by elevation, while aspect or radiation poses orthogonal controls to the elevation. Geology and bedrock are also correlated with elevation, such that hillslopes dominated by crystalline rocks and hydrothermally altered shales and sandstones are primarily found at higher elevations. The plant types are associated with both elevation and radiation, such that the dominant type changes from grassland, aspen, conifer, and grassland to barren from low to high elevations, and conifers are dominant in the north-facing hillslopes.

Instead of relying on a single clustering method, we compared three commonly used clustering methods: hierarchical clustering, k-means, and Gaussian mixture models. Although clustering has been applied in many hydrology or environmental science applications, the selection of methods is often arbitrary and subjective (e.g., Sawicz et al., 2011; Wainwright et al., 2015; Devadoss et al., 2020). Such methodological or conceptual model uncertainties are important to characterize, since they are often larger than parameter uncertainty (Neumann et al., 2003; Ye et al., 2010). Based on the frequency map, we can also identify the hillslopes that are consistently categorized in each zone as most representative hillslopes in that zone. In addition, our results show that different methods yield similar zones provided the distance metrics are the same. Differences could be explained as the further division of an identified zone, such that one method divided a zone into two, due to a particular metric. For example, the conifer-dominated hillslopes were divided into Zone 3, 4 and 7, which could provide insights on the impact of different bedrock types, and topographic features such as slope.

The identified zones are evaluated with the metrics associated with key watershed functions, including foresummer drought, sensitivity of plant productivity, and river nitrogen export from subcatchments. Drought sensitivity is mainly affected by the dominant plant species in the hillslopes, which is consistent with Wainwright et al. (2020). The conifer-dominated hillslopes have lower sensitivity to the droughts, since they are located mostly in north-facing hillslopes and conifer trees have deeper roots. However, there is a difference in two grassland-dominated zones: Zone 5 and Zone 1, with the high elevation grassland

660 less sensitive to droughts, possibly owing to increased water availability in higher elevation. In addition, in a separate study by Yan et al. (2021), the soil thickness and associated parameters such as soil diffusion coefficients are found to be distinctly different between the two zones in this domain (Zone 1 and Zone 2).

Nitrogen export is known to be controlled by multiple factors, such as plant retention (Aber et al. 1998), soil cover (Sickman
665 et al., 2002), river corridor features (e.g., floodplains) (Rogers et al. 2021), and geology (Wan et al., 2019; Arora et al. 2020; Maavara et al. 2021). Our analysis showed that the conifer-dominated zone (Zone 3, 4 and 7) is associated with low nitrogen export, while Zone 2, which has the highest fraction of aspen forests, is associated with higher N export. Conifer forests have previously been observed to have high nitrogen retention (Abrahamsen and Stuanes, 1998, Sollins et al., 1981), particularly through ectomycorrhizal uptake (Höberg et al., 2011). By contrast, deciduous aspen forests produce annual leaf litters, with a
670 low C:N content, releasing nitrogen back to the soil (e.g., DeByle et al., 1985; Köchy and Wilson, 1997, Buck & St. Clair, 2012). In addition, Wan et al. (2019) found that shale was a significant geogenic source of nitrogen, based on the intensive observation site located in Zone 2 of this watershed, which is consistent with the global data synthesis (Houlten et al., 2018). We found that the conifer-dominant hillslopes, including those in Zone 3, were in sandstone or igneous rock regions, or in more resistive shale with a lower fracture density. These factors may be combined to reduce the nitrogen delivery to the river
675 in the conifer-dominated hillslopes, and thus the lower observed river nitrogen exports from that sub-watershed. We acknowledge that our approach does not explicitly account for microtopography or the small-scale heterogeneity in the wetland areas, which are considered biogeochemical hotspots because they are small in area but have outsized impacts on N export (Duncan et al., 2013; Rogers et al. 2021). However, we assume that zonation can still capture a large-scale pattern such that Zone 1 has the largest coverage of the wetland region.
680

5 Conclusion

In this paper, we have developed a watershed zonation approach for characterizing watershed organization and functions based on the bedrock-to-canopy remote sensing (LiDAR, hyperspectral, and electromagnetic surveys) and spatial data layers (soil, geology, and landcover maps). We first delineated the domain into a set of hillslopes, by computing flow routing based on the
685 LiDAR-based DEM. To choose an appropriate scale of hillslopes, we defined the ratio between the across-hillslope variance (i.e., the variance of hillslope-averaged properties) and the overall variance, and investigated the impact of different threshold drainage areas on the ability of the hillslope-averaged metrics to represent the watershed scale heterogeneity. We then applied unsupervised clustering to the hillslope features, including the hillslope-averaged metrics of spatial data layers, as well as the percent coverage of categorical variables within each hillslope. Our major findings are that (1) we can define the scale of
690 hillslopes at which the hillslope-averaged metrics can capture the majority of the overall heterogeneity, particularly for elevation and net annual potential radiation, (2) the identified zones (i.e., the groups of hillslopes) have representative

17

characteristics with respect to co-varied bedrock-to-canopy properties, (3) different clustering methods generate similar zonation *patterns*, given the same distance criteria, and (4) the identified zones could provide information about zone-based watershed functions, including foresummer drought sensitivity and river nitrogen exports.

705 Our approach is similar to catchment clustering and classification of past studies (e.g., Wagener et al., 2007; Krause et al. 2014). Those studies, however, defined the classes of watersheds based on streamflow signatures, which were sparse in space, particularly in headwater catchments. In our study, the expansion of spatial data layers from various remote-sensing data layers provides alternative opportunities to apply clustering to the regions without streamflow measurements. In addition, although our approach is similar the Hydrological Response Unit approach (Flugel, 1997) or hillslope clustering done by Chaney et al.
710 (2018), their main purpose was to parameterize hydrological models. Our analysis confirmed the significance of zonation using the key watershed properties and functions based on observational datasets including novel datasets such as subsurface structures and signatures from airborne EM.

We recognize that every hillslope is unique, each with distinct topographic positions, geology, and vegetation. However, since
715 hillslope-scale studies and experiments are an integral part of watershed science and critical zone science (Hinckley et al., 2014; Tokunaga et al. 2019; Wan et al. 2019), it is important to identify the similarities and differences among the hillslopes. Our approach provides an objective way to classify different hillslopes; including the evaluation of hillslope scales, and the variability associated with different clustering methods. Using our zonation approach, we can guide the experimental plot placement as well as evaluate the representativeness of the selected hillslopes within the domain of interest. After a particular
720 hillslope is identified, we can use pixel-by-pixel clustering to map the heterogeneity within each hillslope associated with microtopography and hillslope positions based on high-resolution images and LiDAR (e.g., Park and Van De Giesen, 2004; Falco et al., 2018; Devadoss et al., 2020; Hermes et al., 2020). This hierarchical representation provides a tractable framework for watershed characterization. Furthermore, Yan et al. (2021) developed two separate parameterizations for the soil evolution model in the two hillslopes that belonged to Zone 1 and 2, and then simulated the variable soil thickness within those hillslopes.
725 In this way, our method can improve model parameterization in large-scale hydrological models by honoring distinct boundaries and water/element export contributions, and provide a new comprehensive way of linking above-and-below ground properties to watershed functions critical to maintaining water resources.

Acknowledgements

18

Deleted: and

Deleted: there are *patterns* in watershed characteristics that can be used to predict zone-based watershed *functions* such as foresummer drought and nitrogen export

Deleted: geology

Deleted: for arebecause they are small in area on N export; Rogers et al. 2021

Moved (insertion) [3]

Deleted: S

Moved up [2]: Our approach provides an objective way to classify different hillslopes by including the variability associated with different clustering methods. Although it is similar to catchment clustering and classification of past studies (e.g., Krause et al. 2014), those studies defined the classes of watersheds based on streamflow signatures, which were sparse; particularly in headwater catchments. The expansion of spatial data layers from various remote-sensing data layers provides alternative opportunities to apply clustering to such spatial features. In addition, although our approach is similar the Hydrological Response Unit approach (Flugel, 1997) or hillslope clustering done by Chaney et al. (2018), their main purpose was to parameterize hydrological models. Our analysis confirmed the significance of zonation using the key watershed properties and functions based on observational datasets. Since hillslope-scale studies and experiments are an integral part of watershed science and critical zone science (Hinckley et al., 2014; Tokunaga et al. 2019; Wan et al. 2019), our zonation approach can

Deleted: Our approach provides an objective way to classify different hillslopes by including the variability associated with different clustering methods. Although it is similar to catchment clustering and classification of past studies (e.g., Krause et al. 2014), those studies defined the classes of watersheds based on streamflow signatures, which were sparse; particularly in headwater catchments. The expansion of spatial data layers from various remote-sensing data layers provides alternative opportunities to apply clustering to such spatial features. In addition, although our approach is similar the Hydrological Response Unit approach (Flugel, 1997) or hillslope clustering done by Chaney et al. (2018), their main purpose was to parameterize hydrological models. Our analysis confirmed the significance of zonation using the key watershed properties and functions based on observational datasets. Since hillslope-scale studies and experiments are an integral part of watershed science and critical zone science (Hinckley et al., 2014; Tokunaga et al. 2019; Wan et al. 2019), our zonation approach can guide the placement of experimental plots more effectively

Moved up [3]: Since hillslope-scale studies and experiments are watershed science and critical zone science (Hinckley et al., 2014; Tokunaga et al. 2019; Wan et al. 2019), our zonation approach can

Deleted: , as well as

Deleted: (Chainy et al., 2019)

Deleted: will

Deleted: ¶

References

Aber, J. D., McDowell, W., Nadelhoffer, K., Magill, A., Berntson, G., Kamakea, M., et al. (1998). Nitrogen saturation in temperate forest ecosystems. BioScience, 48(11), 921–934. https://doi.org/10.2307/1313296

Abrahamsen, G., & Stuanes, A. O. (1998). Retention and leaching of N in Norwegian coniferous forests. Nutrient cycling in agroecosystems, 52(2-3), 171-178.

Anderson, B. T., McNamara, J. P., Marshall, H. P., & Flores, A. N. (2014). Insights into the physical processes controlling correlations between snow distribution and terrain properties, Water Resources Research, 50(6), 4545-4563.

Arora, B., Burrus, M., Newcomer, M. E., Steefel, C. I., Carroll, R. W. H., Dwivedi, D., et al. (2020). Differential C-Q Analysis: A New Approach to Inferring Lateral Transport and Hydrologic Transients Within Multiple Reaches of a Mountainous Headwater Catchment. Frontiers in Water, 2, 24. https://doi.org/10.3389/frwa.2020.00024

Asner, G. P., Martin, R. E., Anderson, C. B., & Knapp, D. E. (2015). Quantifying forest canopy traits: Imaging spectroscopy versus field survey. Remote Sensing of Environment. doi:10.1016/j.rse.2014.11.011

Aytaç, E. (2020). Unsupervised learning approach in defining the similarity of catchments: Hydrological response unit based k-means clustering, a demonstration on Western Black Sea Region of Turkey. International Soil and Water Conservation Research, 8(3), 321-331.

Band, L. E. (1989). Spatial aggregation of complex terrain. Geographical Analysis, 21(4), 279-293.

Band, L. E., Peterson, D. L., Running, S. W., Coughlan, J., Lammers, R., Dungan, J., & Nemani, R. (1991). Forest ecosystem processes at the watershed scale: basis for distributed simulation. Ecological Modelling, 56, 171-196.

825 Band, L. E., Patterson, P., Nemani, R., & Running, S. W. (1993). Forest ecosystem processes at the watershed scale: incorporating hillslope hydrology. Agricultural and Forest Meteorology, 63(1-2), 93-126.

Ball, L. B., Davis, T. A., Minsley, B. J., Gillespie, J. M., & Landon, M. K. (2020), Probabilistic categorical groundwater salinity mapping from airborne electromagnetic data adjacent to California's Lost Hills and Belridge oil fields. Water Resources Research, e2019WR026273.

830

Barfod, A. A., Møller, I., Christiansen, A. V., Høyer, A. S., Hoffimann, J., Straubhaar, J., & Caers, J. (2018). Hydrostratigraphic modeling using multiple-point statistics and airborne transient electromagnetic methods. Hydrology & Earth System Sciences, 22(6).

835 Brantley, S. L., McDowell, W. H., Dietrich, W. E., White, T. S., Kumar, P., Anderson, S. P., ... & Gaillardet, J. (2017). Designing a network of critical zone observatories to explore the living skin of the terrestrial Earth. Earth Surface Dynamics, 5(4), 841-860.

Buck, J. R., & Clair, S. B. S. (2012). Aspen increase soil moisture, nutrients, organic matter and respiration in Rocky Mountain 840 forest communities. *PLoS One*, 7(12), e52369.

Burke, A. R., & Kasahara, T. (2011). Subsurface lateral flow generation in aspen and conifer-dominated hillslopes of a first order catchment in northern Utah. Hydrological Processes, 25(9), 1407-1417.

845 Burt, T. P., & Pinay, G. (2005). Linking hydrology and biogeochemistry in complex landscapes. Progress in Physical geography, 29(3), 297-316.

Carroll, R. W., Bearup, L. A., Brown, W., Dong, W., Bill, M., Williams, K. H. (2018). Factors Controlling Seasonal Groundwater and Solute Flux from Snow-Dominated Basins, Hydrological Processes, 1-16, 10.1002/hyp.12151

850

Chadwick, K. D., Brodrick, P. G., Grant, K., Goulden, T., Henderson, A., Falco, N., ... & Brodie, E. L. (2020). Integrating airborne remote sensing and field campaigns for ecology and Earth system science. Methods in Ecology and Evolution.

Chaney, N. W., Wood, E. F., McBratney, A. B., Hempel, J. W., Nauman, T. W., Brungard, C. W., & Odgers, N. P. (2016). 855 POLARIS: A 30-meter probabilistic soil series map of the contiguous United States. Geoderma, 274, 54-67.

20

Chaney, N. W., Minasny, B.,Herman, J. D., Nauman, T. W., Brungard, C., Morgan, C. L. S., et al. (2019). POLARIS soil properties: 30-m probabilistic maps of soil properties over the contiguous United States. Water Resources Research, 55, 2916–2938. https://doi.org/10.1029/2018WR022797

Chaney, N. W., Van Huijgevoort, M. H. J., Shevliakova, E., Malyshev, S., Milly, P. C. D., Gauthier, P. P. G., and Sulman, B. N.: Harnessing big data to rethink land heterogeneity in Earth system models, Hydrol. Earth Syst. Sci., 22, 3311–3330, https://doi.org/10.5194/hess-22-3311-2018

Clarke, B. A., & Burbank, D. W. (2010). Bedrock fracturing, threshold hillslopes, and limits to the magnitude of bedrock landslides. Earth and Planetary Science Letters, 297(3-4), 577-586.

Colombo, R., Meroni, M., Marchesi, A., Busetto, L., Rossini, M., Giardino, C., & Panigada, C. (2008). Estimation of leaf and canopy water content in poplar plantations by means of hyperspectral indices and inverse modeling. Remote sensing of environment, 112(4), 1820-1834.

DeByle, N. V. (1985). Water and watershed. In: DeByle, Norbert V.; Winokur, Robert P., editors. Aspen: Ecology and management in the western United States. USDA Forest Service General Technical Report RM-119. Rocky Mountain Forest and Range Experiment Station, Fort Collins, Colo. p. 153-160, 119.

Devadoss, J., Falco, N., Dafflon, B., Wu, Y., Franklin, M., Hermes, A., ... & Wainwright, H. (2020). Remote Sensing-Informed Zonation for Understanding Snow, Plant and Soil Moisture Dynamics within a Mountain Ecosystem. Remote Sensing, 12(17), 2733.

Duncan, J. M., Groffman, P. M., & Band, L. E. (2013). Towards closing the watershed nitrogen budget: Spatial and temporal scaling of denitrification. Journal of Geophysical Research: Biogeosciences, 118(3), 1105-1119.

Engstrom, R., Hope, A., Kwon, H., Stow, D., & Zamolodchikov, D. (2005). Spatial distribution of near surface soil moisture and its relationship to microtopography in the Alaskan Arctic coastal plain. Hydrology Research, 36(3), 219-234.

Falco, N., Wainwright, H., Dafflon, B., Léger, E., Peterson, J., Stelzer, H., ... & Hubbard, S. S. (2019). Remote sensing and geophysical characterization of a floodplain-hillslope system in the East River watershed, Colorado. Watershed Functionality Scientific Focus Area.

21

890 Fan, Y., Clark, M., Lawrence, D. M., Swenson, S., Band, L. E., Brantley, S. L., et al. (2019). Hillslope hydrology in global change research and Earth system modeling. Water Resources Research, 55, 1737– 1772. https://doi.org/10.1029/2018WR023903.

Feilhauer, H., Asner, G. P., & Martin, R. E. (2015). Multi-method ensemble selection of spectral bands related to leaf
895 biochemistry. Remote Sensing of Environment. doi:10.1016/j.rse.2015.03.033.

Flugel, W. A. (1997). Combining GIS with regional hydrological modelling usinghydrological response units (HRUs): An application from Germany.Mathematicsand Computers in Simulation, 43(3e6), 297e304.https://doi.org/10.1016/S0378-4754(97)00013-X.

900

Foster, L. M., Williams, K. H., & Maxwell, R. M. (2020). Resolution matters when modeling climate change in headwaters of the Colorado River. Environmental Research Letters, 15(10), 104031.

Gaskill, D.L., F.E. Mutschler, J.H. Kramer, J.A. Thomas, and S.G. Zahony. 1991. USGS Geologic map of the Gothic
905 Quadrangle, Gunnison County, Colorado.

Gillin, C. P., Bailey, S. W., McGuire, K. J., & Gannon, J. P. (2015). Mapping of hydropedologic spatial patterns in a steep headwater catchment. Soil Science Society of America Journal, 79(2), 440. https://doi.org/10.2136/sssaj2014.05.0189.

910 Green, G.N., 1992, The Digital Geologic Map of Colorado in ARC/INFO Format: U.S. Geological Survey Open-File Report 92-0507, 9 p.; http://pubs.usgs.gov/of/1992/ofr-92-0507/ .

Groffman, P. M., Butterbach-Bahl, K., Fulweiler, R. W., Gold, A. J., Morse, J. L., Stander, E. K., ... & Vidon, P. (2009). Challenges to incorporating spatially and temporally explicit phenomena (hotspots and hot moments) in denitrification models.
915 Biogeochemistry, 93(1), 49-77.

Hastie T., Tibshirani, R., and Friedman, J. H. (2001), The elements of statistical learning: data mining, inference, and prediction, Springer, New York, USA.

920 Hermes, A. L., Wainwright, H. M., Wigmore, O. H., Falco, N., Molotch, N., & Hinckley, E. L. S. (2020). From patch to catchment: A statistical framework to identify and map soil moisture patterns across complex alpine terrain. Frontiers in Water, 2, 48.

22

925 Hinckley, E.-L. S., R. T. Barnes, S. P. Anderson, M. W. Williams, and S. M. Bernasconi (2014), Nitrogen retention and transport differ by hillslope aspect at the rain-snow transition of the Colorado Front Range, J. Geophys. Res. Biogeosci., 119, 1281–1296, doi:10.1002/ 2013JG002588.

Högberg, P., Johannisson, C., Yarwood, S., Callesen, I., Näsholm, T., Myrold, D. D., & Högberg, M. N. (2011). Recovery of ectomycorrhiza after 'nitrogen saturation' of a conifer forest. New Phytologist, 189(2), 515-525.
930

Houlton, B. Z., Morford, S. L., & Dahlgren, R. A. (2018). Convergent evidence for widespread rock nitrogen sources in Earth's surface environment. Science, 360(6384), 58-62.

Hubbard, S. S., Gangodagamage, C., Dafflon, B., Wainwright, H., Peterson, J., Gusmeroli, A., ... & Tweedie, C. (2013).
935 Quantifying and relating land-surface and subsurface variability in permafrost environments using LiDAR and surface geophysical datasets. Hydrogeology Journal, 21(1), 149-169.

Hubbard, S. S., Williams, K. H., Agarwal, D., Banfield, J., Beller, H., Bouskill, N., ... & Falco, N. (2018). The East River, Colorado, Watershed: A mountainous community testbed for improving predictive understanding of multiscale hydrological–
940 biogeochemical dynamics. Vadose Zone Journal, 17(1), 1-25.

Hubbard, S. S., Varadharajan, C., Wu, Y., Wainwright, H., & Dwivedi, D. (2020). Emerging technologies and radical collaboration to advance predictive understanding of watershed hydrobiogeochemistry. Hydrological Processes, 34(15), 3175-3182.
945

Jencso, K. G., McGlynn, B. L., Gooseff, M. N., Wondzell, S. M., Bencala, K. E., & Marshall, L. A. (2009). Hydrologic connectivity between landscapes and streams: Transferring reach-and plot-scale understanding to the catchment scale. Water Resources Research, 45(4).

950 Kassambara, A. (2017). Practical guide to cluster analysis in R: Unsupervised machine learning (Vol. 1). Sthda.

Krause, S., J. Freer, D. Hannah, N. Howden, T. Wagener, and F. Worrall (2014). Catchment similarity concepts for understanding dynamic biogeochemical behavior of river basins, Hydrol Process, 28(3), 1554–1560, doi:10.1002/hyp.10093.

955 Köchy, M., & Wilson, S. D. (1997). Litter decomposition and nitrogen dynamics in aspen forest and mixed-grass prairie. Ecology, 78(3), 732-739.

23

Kuentz, A., Arheimer, B., Hundecha, Y., & Wagener, T. (2017). Understanding hydrologic variability across Europe through catchment classification. Hydrology and Earth System Sciences, 21(6), 2863-2879.

Li, L., Maher, K., Navarre-Sitchler, A., Druhan, J., Meile, C., Lawrence, C., ... & Jin, L. (2017). Expanding the role of reactive transport models in critical zone processes. Earth-science reviews, 165, 280-301.

Lintern, Webb, Ryu, Liu, Waters, Leahy, Bende-Michl, and Western (2018), What Are the Key Catchment Characteristics Affecting Spatial Differences in Riverine Water Quality?, Water Resources Research, doi:10.1029/2017WR022172.

Maavara, T., Siirila-Woodburn, E. R., Maina, F., Maxwell, R. M., Sample, J. E., Chadwick, K. D., et al. (2021). Modeling geogenic and atmospheric nitrogen through the East River Watershed, Colorado Rocky Mountains. *PLOS ONE*, *16*(3), e0247907. https://doi.org/10.1371/journal.pone.0247907

Maina, F. Z., & Siirila-Woodburn, E. R. (2020). Watersheds dynamics following wildfires: Nonlinear feedbacks and implications on hydrologic responses. Hydrological Processes, 34(1), 33-50.

Madritch, M. D., Kingdon, C. C., Singh, A., Mock, K. E., Lindroth, R. L., & Townsend, P. A. (2014). Imaging spectroscopy links aspen genotype with below-ground processes at landscape scales. Philosophical Transactions of the Royal Society B: Biological Sciences, 369(1643), 20130194.

Maxwell, R. M. and Kollet, S. J. (2008). Quantifying the effects of three-dimensional subsurface heterogeneity on Hortonian runoff processes using a coupled numerical, stochastic approach, Adv. Water Resour., 31, 807–817, https://doi.org/10.1016/j.advwatres.2008.01.020.

Minsley, B.J. and L.B. Ball (2018), Airborne geophysical characterization of geologic structure in a mountain headwater system, 7th International Workshop on Airborne Electromagnetics, Kolding, Denmark, 17-20 June.

McDonnell, J. J., & Woods, R. (2004). On the need for catchment classification. Journal of Hydrology, 299, 2-3.

McDonnell et al. (2007). Moving beyond heterogeneity and process complexity: A new vision for watershed hydrology, Water Resources Research, 43(7), doi:10.1029/2006wr005467.

24

Mohanty, B. P., Famiglietti, J. S., & Skaggs, T. H. (2000). Evolution of soil moisture spatial structure in a mixed vegetation pixel during the Southern Great Plains 1997 (SGP97) Hydrology Experiment. Water Resources Research, 36(12).

Montgomery, D. R. (2001). Slope distributions, threshold hillslopes, and steady-state topography. American Journal of science, 301(4-5), 432-454.

Neuman, S. P., & Di Federico, V. (2003). Multifaceted nature of hydrogeologic scaling and its interpretation. Reviews of Geophysics, 41(3).

Neuman, S. P. (2003). Maximum likelihood Bayesian averaging of uncertain model predictions. Stochastic Environmental Research and Risk Assessment, 17(5), 291-305.

Newcomer, M. E., Bouskill, N. J., Wainwright, H., Maavara, T., Arora, B., Siirila-Woodburn, E. R., et al. (2021). Hysteresis Patterns of Watershed Nitrogen Retention and Loss over the past 50 years in United States Hydrological Basins. *Global Biogeochemical Cycles*. https://doi.org/10.1029/2020GB006777

Noël, P., Rousseau, A. N., Paniconi, C., & Nadeau, D. F. (2014). Algorithm for delineating and extracting hillslopes and hillslope width functions from gridded elevation data. Journal of Hydrologic Engineering, 19(2), 366-374.

Painter, T. H., Berisford, D. F., Boardman, J. W., Bormann, K. J., Deems, J. S., Gehrke, F., ... & Mattmann, C. (2016). The Airborne Snow Observatory: Fusion of scanning lidar, imaging spectrometer, and physically-based modeling for mapping snow water equivalent and snow albedo. Remote Sensing of Environment, 184, 139-152.

Park, S. J., & Van De Giesen, N. (2004). Soil–landscape delineation to define spatial sampling domains for hillslope hydrology. Journal of Hydrology, 295(1-4), 28-46.

Patton, N. R., Lohse, K. A., Godsey, S. E., Crosby, B. T., & Seyfried, M. S. (2018). Predicting soil thickness on soil mantled hillslopes. Nature communications, 9(1), 1-10.

Peters-Lidard, C. D., Clark, M., Samaniego, L., Verhoest, N. E., Van Emmerik, T., Uijlenhoet, R., ... & Woods, R. (2017). Scaling, similarity, and the fourth paradigm for hydrology. Hydrology and earth system sciences, 21(7), 3701.

de Pasquale, G., Linde, N., & Greenwood, A. (2019). Joint probabilistic inversion of DC resistivity and seismic refraction data applied to bedrock/regolith interface delineation. Journal of Applied Geophysics, 170, 103839.

Pelletier, J. et al. (2018), Which way do you lean? Using slope aspect variations to understand Critical Zone processes and feedbacks, Earth Surface Processes and Landforms, 43(5), 1133–1154, doi:10.1002/esp.4306.

Peng, J., A. Loew, O. Merlin, and N. E. C. Verhoest (2017), A review ofspatial downscaling of satellite remotelysensed soil moisture, Rev. Geophys.,55, 341–366, doi:10.1002/2016RG000543

Peters-Lidard, C. D., Clark, M., Samaniego, L., Verhoest, N. E., Emmerik, T. V., Uijlenhoet, R., ... & Woods, R. (2017). Scaling, similarity, and the fourth paradigm for hydrology. Hydrology and earth system sciences, 21(7), 3701-3713.

Prancevic, J., and J. Kirchner (2019), Topographic Controls on the Extension and Retraction of Flowing Streams, Geophys Res Lett, 46(4), 2084–2092, doi:10.1029/2018gl081799.

Rempe, D. M., & Dietrich, W. E. (2018). Direct observations of rock moisture, a hidden component of the hydrologic cycle. Proceedings of the National Academy of Sciences, 115(11), 2664-2669.

Rodriguez, M. Z., Comin, C. H., Casanova, D., Bruno, O. M., Amancio, D. R., Costa, L. D. F., & Rodrigues, F. A. (2019). Clustering algorithms: A comparative approach. *PloS one*, *14*(1), e0210236.

Rogers, D. B., Newcomer, M. E., Raberg, J. H., Dwivedi, D., Steefel, C., Bouskill, N., et al. (2021). Modeling the Impact of Riparian Hollows on River Corridor Nitrogen Exports. *Frontiers in Water*, *3*, 590314. https://doi.org/10.3389/frwa.2021.590314

Sawicz, K., T. Wagener, M. Sivapalan, P. Troch, and G. Carrillo (2010), Catchment classification: Empirical analysis of hydrologic similarity based on catchment function, Hydrol. Earth Syst. Sci. Discuss., 8, 4495–4534.

Schwanghart, W., Scherler, D. (2014): TopoToolbox 2 – MATLAB-based software for topographic analysis and modeling in Earth surface sciences. Earth Surface Dynamics, 2, 1-7. [DOI: 10.5194/esurf-2-1-2014]

Selby, M.J., (1987), Rock slopes, *in* Anderson, M. G., and Richards, K. S., Slope Stability, Chichester, John Wiley & Sons, p. 475-504.

Sickman, J. O., Melack, J. M., & Stoddard, J. L. (2002). Regional analysis of inorganic nitrogen yield and retention in high-elevation ecosystems of the Sierra Nevada and Rocky Mountains. In *The Nitrogen Cycle at Regional to Global Scales* (pp. 341-374). Springer, Dordrecht.

Sivapalan, M. (2006). Pattern, process and function: elements of a unified theory of hydrology at the catchment scale. *Encyclopedia of hydrological sciences*.

Sollins, P., & McCorison, F. M. (1981). Nitrogen and carbon solution chemistry of an old growth coniferous forest watershed before and after cutting. *Water Resources Research*, *17*(5), 1409-1418.

St. Clair, J. S., Moon, S., Holbrook, W. S., Perron, J. T., Riebe, C. S., Martel, S. J., ... & Richter, D. D. (2015). Geophysical imaging reveals topographic stress control of bedrock weathering. Science, 350(6260), 534-538.

Thompson, S. E., C. J. Harman, P. A. Troch, P. D. Brooks, and M. Sivapalan (2011), Spatial scale dependence of ecohydrologically mediated water balance partitioning: A synthesis framework for catchment ecohydrology, Water Resour. Res., 47, W00J03, doi:10.1029/2010WR009998.

Tokunaga, T. K., Wan, J., Williams, K. H., Brown, W., Henderson, A., Kim, Y., ... & Dong, W. (2019). Depth-and time-resolved distributions of snowmelt-driven hillslope subsurface flow and transport and their contributions to surface waters. Water Resources Research, 55(11), 9474-9499.

Uhlemann, S., B. Dafflon, H.M. Wainwright, K.H. Williams, B. Minsley, K. Zamudio, B. Carr, N. Falco, C. Ulrich, S. Hubbard, Shale bedrock variability correlates with surface morphology and vegetation distribution, and controls hydraulic properties, submitted to Scientific Advances.

Yan, Q., Wainwright, H., Dafflon, B., Uhlemann, S., Steefel, C. I., Falco, N., ... & Hubbard, S. S. (2021). Hybrid data-model-based mapping of soil thickness in a mountainous watershed. Earth Surface Dynamics Discussions, 1-23.

Ye, M., Meyer, P. D., Lin, Y. F., & Neuman, S. P. (2010). Quantification of model uncertainty in environmental modeling. Stochastic Environmental Research and Risk Assessment, 24(6), 807-808.

Wagener, T., M. Sivapalan, P. Troch, and R. Woods (2007), Catchment Classification and Hydrologic Similarity, Geogr Compass, 1(4), 901–931, doi:10.1111/j.1749-8198.2007.00039.x.

Wainwright, H. M., Dafflon, B., Smith, L. J., Hahn, M. S., Curtis, J. B., Wu, Y., ... & Hubbard, S. S. (2015). Identifying multiscale zonation and assessing the relative importance of polygon geomorphology on carbon fluxes in an Arctic tundra ecosystem. Journal of Geophysical Research: Biogeosciences, 120(4), 788-808.

27

100

Wainwright, H. M., Steefel, C., Trutner, S. D., Henderson, A. N., Nikolopoulos, E. I., Wilmer, C. F., ... & Steltzer, H. (2020). Satellite-derived foresummer drought sensitivity of plant productivity in Rocky Mountain headwater catchments: spatial heterogeneity and geological-geomorphological control. Environmental Research Letters, 15(8), 084018.

105
Wan, J., Tokunaga, T. K., Williams, K. H., Dong, W., Brown, W., Henderson, A. N., ... & Hubbard, S. S. (2019). Predicting sedimentary bedrock subsurface weathering fronts and weathering rates. Scientific reports, 9(1), 1-10.

Wood, E. F., Roundy, J. K., Troy, T. J., Van Beek, L. P. H., Bierkens, M. F., Blyth, E., ... & Gochis, D. (2011). Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth's terrestrial water. Water Resources Research,
110   47(5).

Zamudio K.D., Minsley B.J., and Ball L.B, 2020, Airborne electromagnetic, magnetic, and radiometric survey, upper East River and surrounding watersheds near Crested Butte, Colorado, 2017: U.S. Geological Survey, https://doi.org/10.5066/P949ZCZ8.

115

Page 9: [1] Deleted	Haruko Wainwright	9/26/21 9:22:00 AM

Page 10: [2] Deleted	Haruko Wainwright	9/12/21 10:20:00 AM

Page 10: [3] Deleted	Haruko Wainwright	9/12/21 5:26:00 PM

Page 10: [4] Deleted	Haruko Wainwright	9/12/21 10:14:00 PM

Page 11: [5] Deleted	Haruko Wainwright	9/12/21 11:18:00 AM

Page 11: [6] Deleted	Haruko Wainwright	9/12/21 11:21:00 AM

Page 11: [7] Deleted	Haruko Wainwright	9/12/21 11:24:00 AM

Page 12: [8] Deleted	Haruko Wainwright	9/12/21 11:34:00 AM

Page 12: [9] Deleted	Haruko Wainwright	9/12/21 5:48:00 PM

Page 13: [10] Deleted	Haruko Wainwright	9/12/21 5:51:00 PM
Page 14: [11] Deleted	Haruko Wainwright	9/12/21 10:39:00 PM

Page 14: [11] Deleted	Haruko Wainwright	9/12/21 10:39:00 PM

Page 14: [11] Deleted	Haruko Wainwright	9/12/21 10:39:00 PM

Page 14: [12] Deleted	Haruko Wainwright	9/12/21 10:43:00 PM

Page 14: [12] Deleted	Haruko Wainwright	9/12/21 10:43:00 PM

Page 14: [13] Deleted	Haruko Wainwright	9/12/21 1:37:00 PM

Page 14: [13] Deleted	Haruko Wainwright	9/12/21 1:37:00 PM

Page 14: [14] Deleted   Haruko Wainwright   9/12/21 10:31:00 PM