WHYCHUS CREEK

SUBSTRATE ANALYSIS

Quantifying Geomorphic Responses of Habitat Restoration Projects: Pilot project

Prepared by:
Alex Scagliotti
Lauren Mork

Prepared for:
Upper Deschutes Watershed Council

Abstract

It is generally accepted in stream ecology that habitat heterogeneity and patchiness at multiple scales increases ecosystem resilience through niche diversification. Heterogeneous stream habitats include a complex mosaic of hydraulic features, large woody debris, anabranches, substrata and channel forms - this complexity tends to increase as streams progress through evolutionary stages. Recent restoration work on Whychus Creek in Central Oregon has sought to create complex, late-stage systems in order to improve the ecological function of artificially simplified reaches. One way to measure and track lotic system habitat complexity is through substrate analyses. The goal of this pilot project was to develop a replicable and robust monitoring protocol that quantifies substrate heterogeneity conditions among four priority reaches. We developed a monitoring protocol that utilizes three methods to capture substrate heterogeneity on four, 500-m reaches of the creek. Each sample reach included a nested sampling design of 12 floodplain-wide transects that allowed me to quantify micro, meso or macro-level substrate heterogeneity. We collected data using standard pebble counts, two-dimensional areal plot estimates and one-dimensional patch width measurements. We used the data from each of these three methods to calculate habitat heterogeneity using four metrics – Simpson's Diversity Index, Shannon's Evenness Index, Lloyd's Index of Patchiness and Fortin's Spatial Diversity Index. The results indicated that the two recently restored reaches were on average, 38% more heterogeneous than the untreated reach while the older, more established project reach was on average, only 15% more heterogeneous. The chi-square test for independence for the pebble count indicated significant differences between all the reaches and substrate classes (X^2 (18, N = 1865) = 210.23, p <.001) except one – which signaled that the untreated reach requires a slightly larger sample size in future years. For the plot method, the differences among the reaches were more significant with X^2 (18, N = 2306) = 836.57, p <.001. The plot method resulted in the highest Cramer's V value of 0.35 ($p < 0.001$) - indicating a strong relationship between substrate composition and individual reach. These results illustrate that the three methods were robust enough represent stream substrate conditions.

Table of contents

Introduction

For over a century, Whychus Creek has undergone largescale alterations to its pre-settlement function. Changes in land management, urbanization, agriculture, recreation and infrastructure have resulted in degraded reaches and simplified stream habitats. Straightened and 'bermed' channels, local extirpation of ecosystem engineers, water diversions, on-channel infrastructure and other influences have contributed to this degradation[1]. Recent work by the Upper Deschutes Watershed Council (UDWC) and restoration partners have contributed to a concerted effort to improve the ecological function of Whychus Creek.

A useful approach to measure, track and compare ecosystem response to restoration efforts as well as large changes within the basin is through a substrate analysis. Substrate heterogeneity and patchiness can be important determinants of lotic system health because they reflect key components of stage zero restoration[2] and are some of the earliest indicators of change within the watershed[3]. A complex mosaic of substrate classes reflects hydraulic and topographic heterogeneity (e.g. riffles, pools, backwaters etc.), allows for ontogenetic migration of aquatic species, increases diversity in the invertebrate community through niche diversification[4] and improves the resiliency of the ecosystem to disturbance[5]. Since one of the primary goals of UDWC restoration projects is to restore simplified reaches into complex, late-evolutionary stage systems (Stage 8 or 0 per Cluer and Thorne's Stream Evolution Model[2]), including a substrate analysis protocol as a tool in the larger monitoring scheme will provide useful information about a key indicator of stream evolution.

The objectives of developing this monitoring protocol were as follows:

1. Develop and test monitoring methods for long-term, replicable analysis of substrate patch heterogeneity of current and future restoration sites along Whychus Creek.

2. Develop and test a variety of metrics that can be used to compare substrate conditions among the four sites.

3. Characterize the substrate patch diversity and spatial distribution in four reaches of Whychus Creek that are in various stages of stream evolution.

---

[1] Mork, L. and R. Houston. 2016. "2015 Whychus Creek monitoring report." Last modified December 2016. http://www.upperdeschuteswatershedcouncil.org/wp-content/uploads/2018/01/2015-Whychus-Creek-Monitoring-Report_FINAL.pdf

[2] Cluer, Brian and Colin Thorne. 2014. "A stream evolution model integrating habitat and ecosystem benefits." *River Research and Applications* 30, no. 2 (Februrary): 135-154. https://doi.org/ 10.1002/rra.2631

[3] Dietrich, William E., James W. Kirchner, Hiroshi Ikeda and Fujiko Iseya. 1989. "Sediment supply and the development of the coarse surface layer in gravel-bedded rivers." *Nature* 340: 215. https://doi.org/10.1038/340215a0

[4] Milesi, Syliva V., Sylvain Doledec and Adriano S. Melo. 2016. "Substrate heterogeneity influences the trait composition of stream insect communities: an experimental in situ study." *Freshwater Science* 35, no. 4 (December): 1321-1329. https://doi.org/ 10.1086/688706

[5] Pederson, Morten and Nikolai Friberg. 2007. "Two lowland stream riffles – linkages between physical habitats and macroinvertebrates across multiple spatial scales." *Aquatic Ecology* 41, no. 3: 475-490. https://doi.org/ 10.1007/s10452-004-1584-x

Comparing the geomorphic response of fluvial processes among sites that represent various stages of stream evolution was an exciting and insightful project. This is because the results of this monitoring program quantified and compared the substrate conditions in the three restoration reaches as well as conditions in a pre-project, untreated reach. The results of this pilot project indeed indicated that there were higher rates of heterogeneity and patchiness in the restoration project reaches than in the untreated reach. Furthermore, each of the three methods of data collection used for this study showed significant differences between reaches and consistent relative values of heterogeneity which indicates that they were responsive to the geomorphic conditions of the creek. Continued annual monitoring using these methods will be able to expand upon the conclusions found here and allow for more in-depth analyses that can signal changes within the basin.

Definitions

The terms *habitat heterogeneity*, *patchiness* and *complexity* are widely used in ecological literature and they are often used interchangeably as catch-all terms. We used Li and Reynolds' definition of heterogeneity which includes two components – the system property and its complexity or variability[6]. In this case, the property measured was substrate patch diversity. The second component is complexity (categorical descriptors) or variability (numerical descriptors). The categorical descriptors in this case were substrate class (per classification system), such as fine sediment, gravels, boulders, etc., while the numerical descriptors are the size and abundance of each class. Therefore, substrate patch heterogeneity will be determined by assessing complexity and variability among reaches. We defined a *patch* as an area where the dominant substrate size class is relatively homogenous and differs from the larger matrix of its surroundings[7]. *Patch boundaries* were considered as an abrupt or gradual discontinuity between substrate classes.

[6] Li, H. and J.F. Reynolds. 1995. "On definition and quantification of heterogeneity." *Oikos* 73, no. 2 (June): 280-284. https://doi.org/10.2307/3545921

[7] Winemiller, Kirk O., Alexander S. Flecker and David J. Hoeinghaus. 2010. "Patch dynamics and environmental heterogeneity in lotic ecosystems." *Journal of the North American Benthological Society* 29, no. 1 (March): 84-99. https://doi.org/ 10.1899/08-048.1

Methods

Substrate monitoring on Whychus Creek was conducted between August 13 and August 26, 2018. Pebble count and plot methods were conducted August 13-15; transect patch surveys were conducted on August 26. This timeline represents base-flow conditions which improved the ease of measurement and reduced the chance that conditions would change between sampling days.

Site layout

Sampling reaches

The four reaches we sampled in this project were Camp Polk Reach 2 (CP-R2), Whychus Canyon Reach 3 (WC-R3), and Whychus Canyon Reach 4 split into upper and lower sections (WC-R4 Forest and WC-R4 Delta). CP-R2 is the older restoration project reach with the last major work ending in 2012, followed by WC-R4 with major work ending in 2016. WC-R3 is an untreated reach. We established one 500-m valley-length sampling reach at CP-R2, one at WC-R3, and two at WC-R4. The last reach was split into two sections to better characterize the variability within this ~1.6-km stream restoration project.

Transect layout

Sampling reaches were approximately 50 to 150 meters wide depending on the width of the active floodplain. We split each 500-m reach into four groups of three transects each (12 transects total) that ran perpendicular to the valley. Each of the three transects within a group were placed 20 meters apart so that each grouping was 40 meters wide. We spaced the middle transects of the four groups 125 meters apart throughout the reach. The transects spanned the width of the active floodplain (figure 1) which we defined as the area that is expected to be inundated on a two-year recurrence interval*. Transects included the active floodplain because the entire area may provide habitat during high flows from late spring runoff. Additionally, given the dynamic nature of late-stage streams, channels may shift, close off, widen or narrow as they evolve and it was important to establish a monitoring protocol that captures those changes.

We designed a grouped layout of nested sampling scales because it us permitted more freedom in the analysis than using equidistant transects[8]. A grouped design allowed us to calculate patch diversity at both the meso-habitat and reach scales. In addition, it was necessary to include a grouped layout when analyzing spatial diversity indices because they rely on the sequence of values between defined and regular spatial intervals. By comparing the combined transect results in each of the four groups within a reach, we increased the sample sizes being compared and therefore the probability that the sampling scheme in each group accurately represented the condition of the stream.

---

[8] Boyero, Luz. 2003. "The quantification of local substrate heterogeneity in streams and its significance for macroinvertebrate assemblages." *Hydrobiologia* 499, no. 1-3 (June): 161-168.

*The floodplain of WC-R3 is hydrologically disconnected from the creek and does not fit our definition based on a two-year recurrence interval. The transects were set in what is anticipated to be the active floodplain after the reach is treated.

Field methods for locating transects

We used aerial-image maps with georeferenced transects to delineate each 500-m reach so that they can be found again with a GPS unit for future monitoring. In addition, we 'monumented' the outside boundaries of each transect with capped rebar and fiberglass rods to ensure that the same area will be measured and results can be accurately compared with future substrate monitoring results. To improve the accuracy of remaining in-line with the transects, the surveyors temporarily placed several pieces of flagging along intermediate points between the outside edges of the sampling area. We used a GPS unit and georeferenced maps of each site to determine the placement of the intermediate flagging along the transects. This ensured that at least two pieces of flagging were always visible to the surveyors which proved to be more accurate and timely than using a leveling line between end-points.

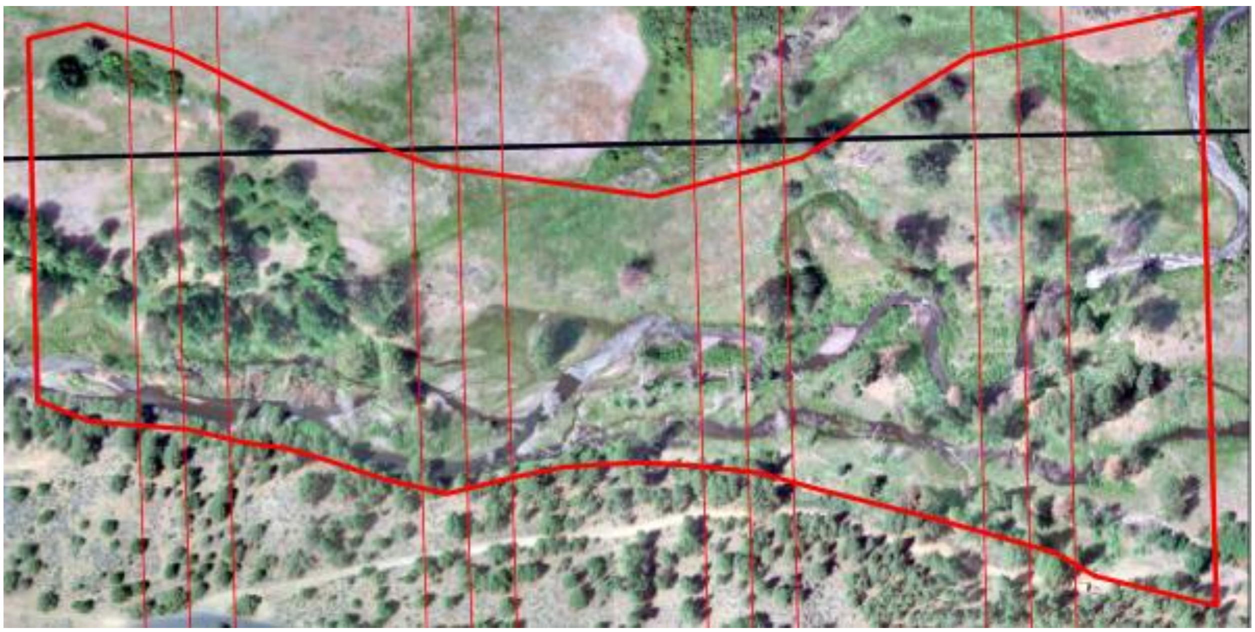


Figure 1: Aerial image map of CP-R2 with transect layout. The black line runs parallel to the valley, the thick red outline delineates the approximate active floodplain boundary of the reach and the thin red lines are the transects that run perpendicular to the valley.

Data collection

Pebble count

Using a point-intersect method from the crosshairs of a 0.5 by 1 meter plot (figure 2), the surveyor selected a pebble sample every 0.5 meters along the entire width of each transect. The surveyor placed the plot lengthwise along the transect (perpendicular to the flow) so that two samples could be selected per placement. The surveyor blindly reached under the two upstream and outermost crosshairs (figure 2) and selected the first particle touched[9]. After two particles

[9] Bunte, Kristin and Steven R Abt. 2001. "Sampling frame for improving pebble count accuracy in coarse gravel-bed streams." *Journal of the American Water Resources Association* 37, no. 4 (August): 1001-1014. https://doi.org/10.1111/j.1752-1688.2001.tb05528.x

were selected, measured, and recorded, the sampling frame was flipped along its shortest axis and two more samples taken. Because of the way we designed the frame, this ensured that one particle was selected every 0.5 meters. We repeated this process along all non-vegetation dominant segments of each transect. The surveyor remained downstream of the plot during all pebble counts so as not to disturb the substrate being sampled. The surveyor measured the intermediate axis (i.e. neither the shortest nor longest axis) of each particle by placing it through the smallest hole it could fit through on a gravel card (figure 3)[10]. Larger cobbles and boulders that could not fit through the largest hole were measured along their intermediate axis with a metric ruler. The surveyor determined small, immeasurable substrate classes by feel – silt was universally marked as 0.5mm when it was smooth between fingers and sand was universally marked as 1mm when it had a gritty feeling. Only mineral substrates were sampled. Once recorded, we later 'binned' the data by size classes. The size classes can be broken up into a simple distribution of broad size classes of silt, sand, gravel, cobble, boulder and bedrock or they can be segregated into finer detail of smaller ranges (e.g. small, medium or large cobble). After preliminary data analysis, we chose the simple distribution as the input data for the diversity indices because 1) it represented a more parametric distribution of size classes, 2) so it could be easily compared with the other two sampling methods implemented in this study, and 3) because biota likely show clearer trends between broad size classes (e.g. gravel, cobble, boulder) than small-range classes (e.g. small, medium or large gravel).

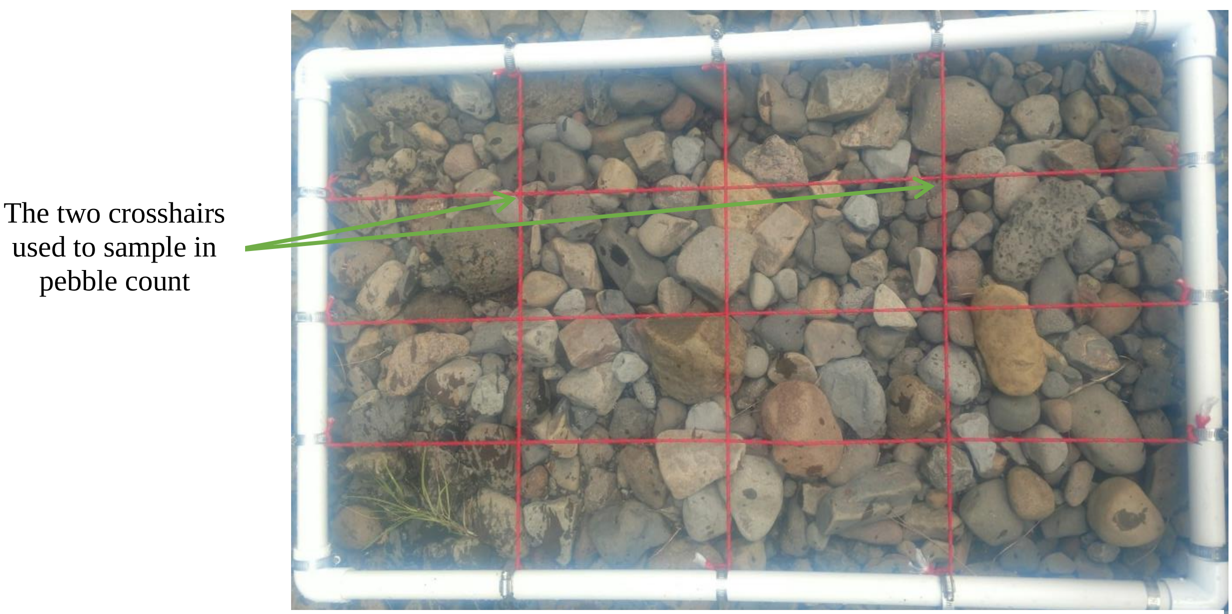


Figure 2: The 1 X 0.5 meter plot used to select particles for the pebble count as well as estimate patch cover in the plot method.

[10] Bunte and Abt 2001

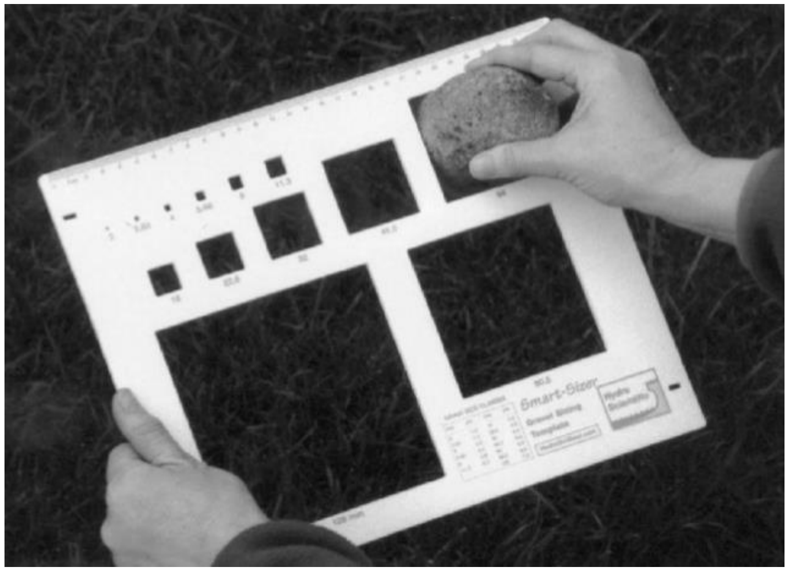


Figure 3: Gravel card with 13 size ranges and a metric ruler for measuring larger substrata.[11]

Plot method

Using the same 0.5 by 1 meter plot (figure 2), the surveyor conducted ocular estimates to determinet total patch area within the plot[13]. Patches were characterized by areas of a dominant size class and patch borders were delineated by distinct changes in the dominant cover of a particular size class. The photo of the plot (figure 2) would be characterized as 100% gravel. Though there are individual particles that would qualify as cobble, there are no distinct cobble patches and the dominant size class present is gravel. We did not count patches that made up less than 10% of the plot area. The 16 quadrants within the plot improved the accuracy of delineating patch areas by counting the number of 12.5 x 25 cm segments that contained the *i*th size class. The patch classes and determinations we used were:[12]

Silt: Smooth between fingers
Sand: Gritty between fingers
Gravel: BB to tennis ball sized
Cobble: Tennis ball to basketball sized
Boulder: Larger than a basketball
Vegetation: Live vegetation including riparian vegetation, macrophytes and periphyton
Detritus: Dead and decaying plant material including significant log jams, debris piles and dead/decaying algae colonies

---

[11] Bunte and Abt 2001

[13] Collins, Scott L. and Melinda D. Smith. 2006. “Scale-dependent interaction of fire and grazing on community heterogeneity in tallgrass prairie.” *Ecology* 87, no. 8 (August): 2058-2067. https://doi.org/ 10.1890/0012-9658(2006)87[2058:SIOFAG]2.0.CO2

[12] U.S. Forest Service 2012. *Stream inventory handbook: level I & II.* U.S. Forest Service. PDF.

Each transect consisted of three semi-randomized plot placements for a total of 36 plot placements per reach. Since each transect spanned the active floodplain and the majority of that area (approximately 70%) was dominated by vegetation, completely randomized placements would have similarly been dominated by vegetation and would not have captured channel diversity. To mitigate this, we assessed all the areas not dominated by vegetation and then used a random number generator that ranged from 0-100 to determine at what percentage of the width of each segment to place a plot. For example, if a transect crossed three channels (flowing or dry) and the random number generator produced 10, 90 and 70, then we placed plots at 10% across the non-vegetation-dominant area of one channel, 90% across another and 70% across the third channel. If there were more or less than three channels, we prioritized larger areas of mixed substrata over narrower ones.

Transect-wide patches

At the middle transect of each group, we used a 60-meter tape to measure the one-dimensional width of each patch class across the active floodplain[13]. We determined the width of each individual patch by marking and measuring the boundaries where the dominant class was replaced by another dominant class. For example, along a given transect, 0 – 5m might be marked as vegetation, 5 - 6.1m as sand, 6.1 – 6.8m as gravel etc. We included all of the same substrate classes in this method as we included in the plot method – silt, sand, gravel, cobble, boulder, vegetation and detritus. We Sampled four transects per reach rather than the 12 transect samples of the other two methods for two reasons. First, the high dominance of vegetation along the reach negatively skewed the diversity indices and therefore, the index outputs could not be accurately compared to the other two methods. Second, this method was the most time consuming of the three, with an estimated four days required to measure all 48 transects, whereas the other two methods required less than three days combined. Since this study was meant to develop a replicable and *timely* monitoring protocol, sampling all 12 transects with this method did not fit within our goals.

[13] Meyer, Kate. 2017. *Deer Creek floodplain restoration project: A Stage 0 Restoration Case Study in Western Oregon*. Report. Portland, OR: U.S. Forest Service, Willamette National Forest. http://www.rrnw.org/wp-content/uploads/3.2FINAL_Deer-Creek-Floodplain-Restoration-Project_RRNW.pdf

Method justification

Each of the three methods used in this study encompassed different aspects of the creek's substrata. The three methods were intended to be used in concert so that the strengths of each strategy would counteract the weaknesses of another (table 1) .

Table 1: Strengths and weaknesses of each method of data collection

Method	**Strengths**	**Weaknesses**
Pebble count		
	Most objective	Assumes results of each transect mirrors general patch makeup
	Point-intersect particle selection & gravel card measurements reduce bias	Some selections are anomalous & not representative of dominant patch class
	Captures conditions of the entire transect	Only captures mineral substrates
	Support a large variety of statistical analyses	Does not capture the condition of the entire transect
Plot method		
	Measures the two-dimensional patch makeup	More subjective than other methods
	Quadrants within the plot improve accuracy & consistency of ocular estimates	Does not capture the condition of the entire transect
	Includes more classes of substrates than pebble count	
	Most time efficient method	
Transect patches		
	Measures large patches	Most time-intensive method
	Captures the condition of entire transect	Measures only one-dimensional patch size
	Includes more classes of substrates than pebble count	Continuous data limits statistical analyses

Statistical analysis

Indices of heterogeneity

To quantify substrate patch heterogeneity, we used several diversity indices applied in other fluvial geomorphological studies[14,15] and other related fields such as landscape ecology[16] and biogeography[17]. Each of these indices can be more or less responsive to changes in the input data so they were meant to be interpreted in the context of each other – in our analysis, any individual indicator of diversity was too simplistic and could not accurately represent the status of the creek[20]. For detailed information on strengths and weaknesses of each indicator, see the Study Proposal.

Simpsons Diversity Index (SDI): $$\sum_{i=1}^{c} \frac{n_i(n_i - 1)}{N(N-1)}$$

Where n is the frequency or area of the ith class and N is the total of all classes. The output range is 0 < SDI < 1, with higher values indicating higher diversity. SDI Responds to number of patches and proportions but is most the responsive to class richness.[18]

Shannon's Evenness Index (SEI): $$\sum_{i=1}^{n} -\frac{p_i \ln p_i}{\ln N}$$

Where P_i is the percent composition of the ith class and N is the total of all classes. The output range is 0 < SDI < 1, with higher values indicating a more even composition of classes and lower values indicating a reach with a few dominant classes. This responds most to changes in proportionality between classes [20].

Lloyd's Index of Patchiness (LIP): $$\frac{(\bar{x} + \sigma^2)/(\bar{x} - 1)}{\bar{x}}$$

Where $\bar{x}$ is the mean frequency of each substrate class between transect groups and σ^2 is the variance of the frequency of each class between transect groups. Rather than providing a single number as an output that represents the entire reach, this index indicates how congregated or

14 Beisel, Jean-Nicolas, Philippe Usseglio-Polatera and Jean-Claude Moreteau. 2000. "The spatial heterogeneity of a river bottom: a key factor determining macroinvertebrate communities." *Hydrobiologia* 422 (April): 163-171. https://doi.org/

15 Boyero, Luz. 2003. "The quantification of local substrate heterogeneity in streams and its significance for macroinvertebrate assemblages." *Hydrobiologia* 499, no. 1-3 (June): 161-168.

16 Collins and Smith 2006.2

17 Fortin, Marie-Josee, Sergej Payette and Kim Marineau. 1999 "Spatial vegetation diversity index along a postfire successional gradient in the northern boreal forest." *Ecoscience* 6, no. 2 (January): 204-213. https://doi.org/10.1080/11956860.1999.11682521

18 Brown, Bryan L. 2003. "Spatial heterogeneity reduces temporal variability in stream insect communities." *Ecology Letters* 6, no. 4 (April): 316-325. https://doi.org/10.1046/j.1461-0248.2003.00431.x

dispersed each substrate class is within the reach. A value >1 indicates dispersed occurrences of a particular class and <1 indicates that a class is congregated is specific regions[19] (Appendix A).

Fortin's Spatial Diversity Index (Spl.DI): $$\sum_{i=1}^{n} -\frac{p_i \log p_i}{\log N}$$

Where P_i is the percent composition of the *i*th class and *N* is the total of all classes. Rather than providing a single value as the output that represents the entire reach, this index indicates how widespread and proportional each class is throughout the reach. The output values are 0 < Spl.DI < 1, with higher values indicating more spatially diverse spread[20] (Appendix B).

Analyzing diversity index values

SDI and SEI scores combine reach-wide proportions of all substrate classes and their proportions for a single value output that represents the whole reach. Conversely, LIP and Spl.DI take into account spatial distribution between group samples and the output is individual substrate class values throughout a reach. For example, a gravel frequency of 11, 3, 20, and 0 between transect groups would result in different values for both LIP and Spl.DI than would 7, 8, 9 and 10 - even though the total abundance and mean are the same for both samples. Neither of those two distributions would affect the overall output values of SEI or SDI.

For Spl.DI and LIP, we measured the differences between samples at the group level (N=4). For this reason, the tables and figures for each reach in our results show the Spl.DI and LIP value for each substrate class as well as a combined mean value. To combine the Spl.DI values of individual substrate classes into a single score, we calculated the mean score among classes to represent overall spatial diversity for the reach. To combine LIP scores into a single reach-wide value, we counted all scores where LIP < 1.0 and divided by the total number of possible substrate classes. For example, CP - R2 had four substrate classes as scoring less than 1.0 under the plot method so the overall output value for the reach was 4/7 = 0.57. Converting these values to a $0 < x < 1$ spectrum allowed for more standardized comparisons between reaches and methods.

[19] Xiao, C.L., J.J. Hao and K.V. Subbarao. 1997. "Case study #2: Lloyd's Index of Patchiness." Accessed October 13, 2018. http://www.apsnet.org/edcenter/advanced/topics/EcologyAndEpidemiologyInR/SpatialAnalysis/Pages/CaseStudy2Lloyd%27sIndexofPatchiness.aspx

[20] Fortin and Marineau 1999.

We determined the a confidence interval (CI) for these values based on formulas that calculate the variance of the scores. However, after considerable research, we found equations for only SDI and SEI – likely because these are widely used and well known indices whereas LIP and Spl.DI have a narrower field of use. This was one of the weaknesses associated with using LIP and Sptl.DI to determine patch heterogeneity. The variance formulas for SDI and SEI are:

$$VAR_{SDI} = \frac{\sum_{i=1}^{s}\left(\frac{n_i}{N}\right)^3 - \left[\sum_{i=1}^{s}\left(\frac{n_i}{N}\right)^2\right]^2}{N/4}$$

$$VAR_{SEI} = \frac{\sum_{i=1}^{s} n_i \ln(n_i)^2 - \left[\sum_{i=1}^{s} n_i \ln(n_i)\right]^2 / N}{N^2}$$

Where s is the number of classes, n_i is the frequency of the ith class and N is the total sample[21].

Testing for significance between reaches and methods

We used a chi-square test of independence of both the plot method and pebble count to test whether reach totals for substrate classes were significantly different between reaches. For the chi-square test of independence, data must be in whole-number integers and not continuous data. For this reason, the areal plot method data was converted to the frequency of occurrence of 12.5 by 25cm quadrants within each plot. The plot had a total of 16 segments of this size for a total sample size of N=576 per reach. We could not test the data from the transect patch method because it included continuous data. While the chi-square test indicates whether the reaches can be said to be significantly different from each other, or more specifically, whether substrate distributions are independent (accepting the null) or dependent (rejecting the null) on a reach, it does not indicate how strong the relationship between these variables is. Cramer's V does test the strength of associations between the reach and substrate composition so we used this test to determine *how* different the reaches were from each other.

---

[21] Statsdirect. 2018. "Diversity of classes." Accessed 15, October 2018. https://www.statsdirect.com/help/nonparametric_methods/diversity.htm

Results and Discussion

Summary of substrate distributions

Pebble count summary and comparison

Pebble count results show that there was more non-vegetated habitat available for aquatic biota in the restoration project reaches than the untreated WC-R3. The largest sample, and therefore greatest non-vegetated area of the active floodplain was WC-R4 Delta with 617 samples taken, followed by WC-R4 Forest with 555 samples, CP- R2 with 519 samples and finally WC-R3 with 174 samples. These values cannot be explicitly compared because the floodplain widths varied slightly between and within reaches but these values nevertheless point to several trends. First and most obvious is that WC-R3 had significantly less mineral and by inference, aquatic habitat available than the other three reaches. Additionally, WC-R3 was dominated by cobble (49%), making it the only reach with this dominant substrate class. This is an expected outcome since a smaller sample size indicates that the flow is confined to a smaller area and has more energy to transport smaller particles. Conversely, the largest majority of the samples in the three project reaches fell in the 'gravel' class with 45% in CP-R2, 41% in WC-R4 Delta and 35% in WC-R4 Forest (figure 4), indicating that the project reaches likely disperse much of the hydraulic energy, resulting in smaller size class distributions.

Comparison of the reaches also indicates that CP-R2 class composition most closely resembled a parametric distribution of size classes while the two WC-R4 sections displayed higher rates of finer material. The distributions in the two WC-R4 sections may begin to mirror CP-R2 in following years as riparian and hydrophytic vegetation have a chance to recolonize areas of finer sediments. CP-R2 has likewise had more time for vegetation to become established which may also indicate why there were less mineral substrata present than the recently completed projects in Whychus Canyon.

CP-R2 shows a relatively even spread of its interquartile range (figure 6) which indicates that the sizes from 25% above and below the median (the 25-75% range of substrata) were relatively evenly distributed throughout the reach. Conversely, the two WC-R4 reaches had smaller median diameters and size distributions skewed towards smaller particles. WC - R4 Forest had the widest interquartile range (1- 45mm) as well as the widest overall range (0.5 to 1,200 mm) of the three project reaches. The WC-R3 size distribution had a median substrate size of 64 mm which qualifies the median size as "cobble" under the simple distribution (table 2) or "very coarse cobble" – "small boulder" under the detailed distribution (table 3).

Table 2: Simple size class distributions – pebble count

Substrate class	Bin (mm)
silt	0.5
sand	1 - 4
gravel	5 - 45
cobble	46 -256
boulder	257 - 1200

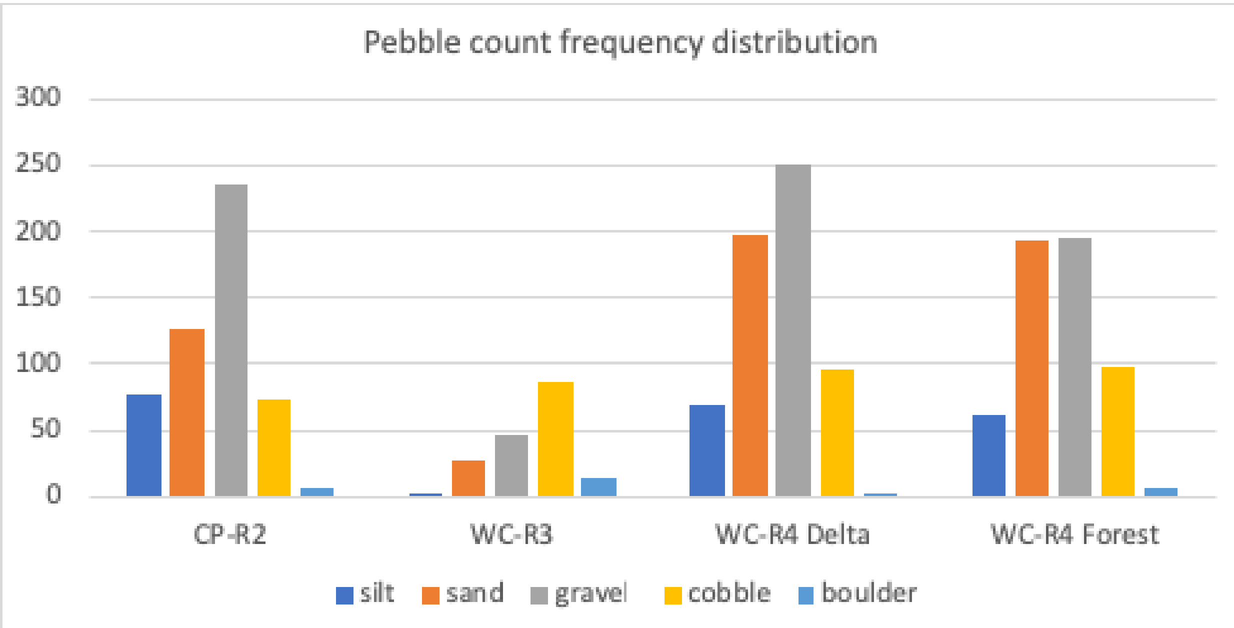


Figure 4: Substrate class distribution comparison between reaches – pebble count

Table 3: Detailed size class distributions – pebble count

Substrate class	Size class	Bin (mm)
clay/silt		0.5
sand		1-2
Gravel	Very fine	3-4
	fine	5-8
	medium	9-16
	coarse	17-32
	very coarse	33-64
Cobble	small	65-128
	large	129-256
Boulder	small	257-512
	large	513-1200

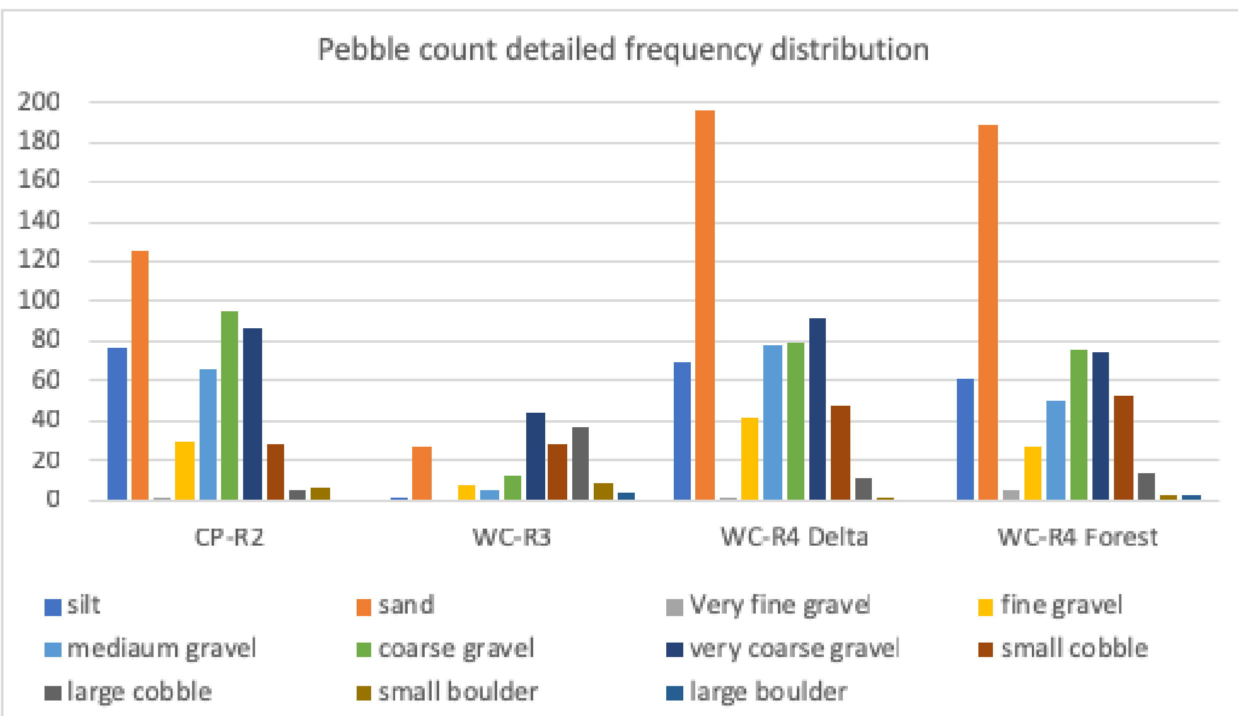


Figure 5: Detailed substrate class distribution comparison between reaches – pebble count

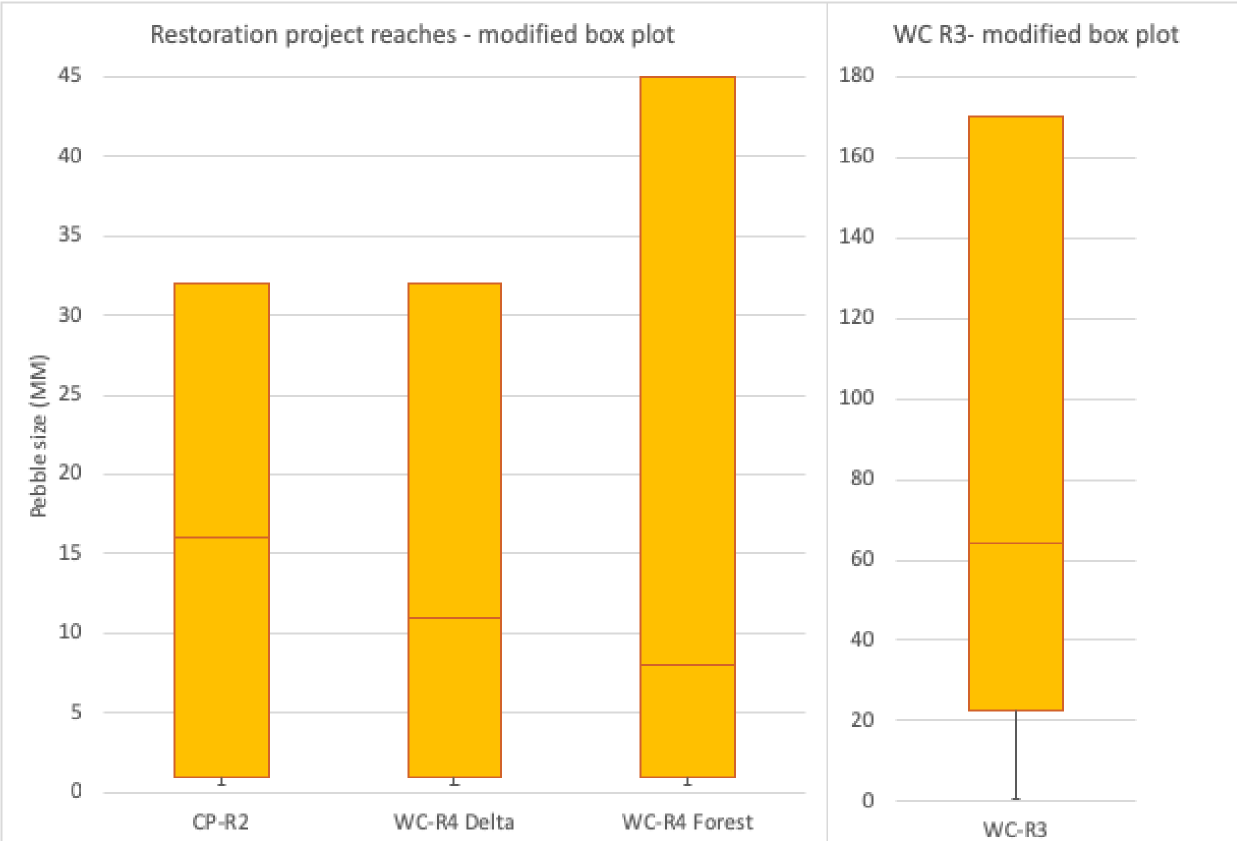


Figure 6: Modified box-whisker plot of mineral substrate diameter distributions between reaches in (mm) This input data was not binned into classes but it displays the gross size distributions with the lowest whisker indicating the minimum size and first quartile, the bottom box indicating the second quartile range, the orange line indicating the median size and the upper box indicating the third quartile range. The fourth quartile cannot be displayed because the maximum diameter found in each reach is orders of magnitude larger than the rest of the distribution (table 4). WC-R3 contained much higher diameters so it is displayed on a separate scale.

Table 4: Substrate intermediate axis diameters (mm)

	CP-R2	**WC-R4 Delta**	**WC-R4 Forest**	**WC-R3**
Minimum	0.5	0.5	0.5	0.5
1st quartile	1	1	1	22.6
Median	16	11	8	64
3rd quartile	32	32	45	170
Maximum	340	310	1200	900

Plot method summary and comparison

Under this method, CP-R2 once again had a strong dominance of gravel (49%) over other substrates while WC-R3 had a strong, though slightly lower dominance of cobble (43%). Similar to the results from the pebble count, both WC-R4 reaches had a more moderated spread of substrate classes (figure 7), indicating that there was a higher diversity of substrata in these reaches than WC-R3 or CP-R2. When comparing only mineral substrates from the plot method to the pebble count, the four reaches had similar distributions (table 5).

Though there are noticeable differences in percent-cover ratios between the two methods, there is only one difference in intra-reach substrate rank between the two methods. Under the pebble count method in WC-R3, boulders were the fourth most abundant substrate (7.5%) and in the plot method, they were the third most abundant (8.1% excluding biotic substrata). Though the percentage is similar, the rank in the pebble count indicates that there was more cover of sand than boulders. This is unsurprising since a pebble count only counts a boulder as a single occurrence in the overall makeup whereas the plot method takes into account its size - a single boulder may dominate an entire plot area. For this reason, the plot method shows that on average there was 45% more coverage of boulders throughout all reaches than the pebble count.

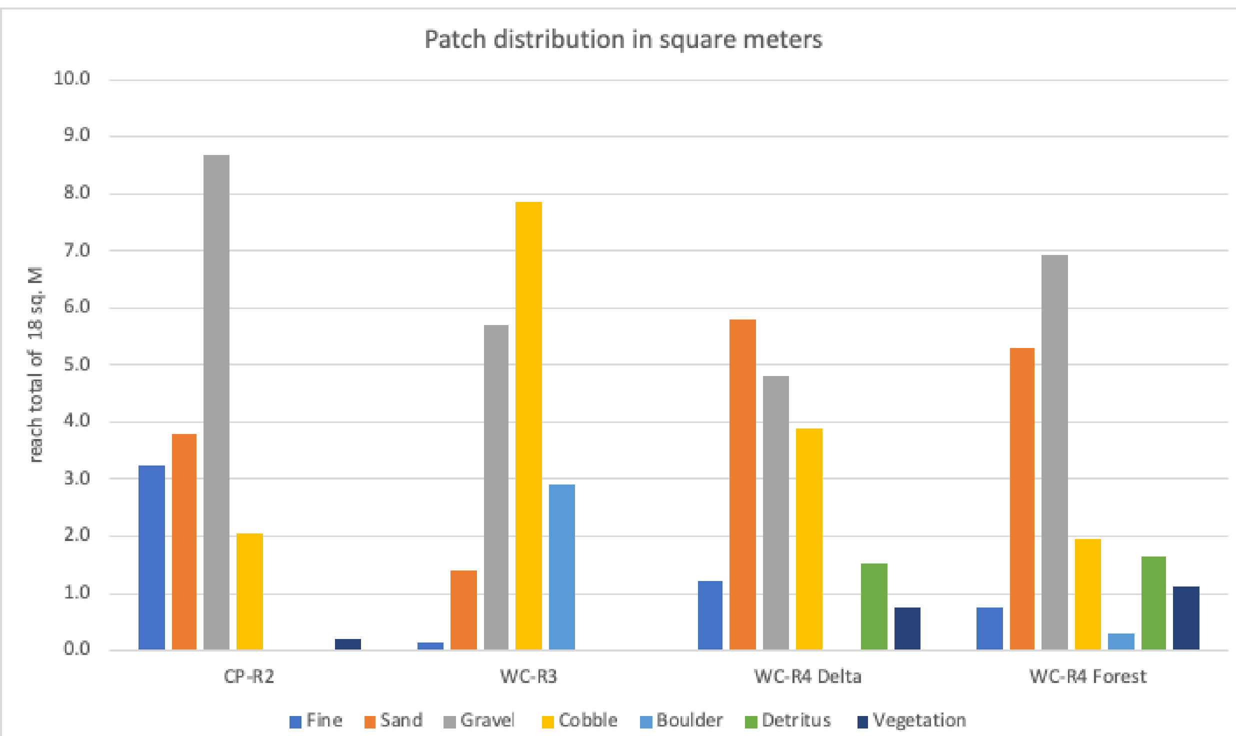


Figure 7: Substrate class distribution between reaches – plot method

Table 5: Mineral substrate distribution comparison between two methods.

		Silt	Sand	Gravel	Cobble	Boulder
Pebble count	CP-R2	14.8%	24.5%	45.5%	14.1%	1.2%
	WC-R3	0.6%	15.5%	27.0%	49.4%	7.5%
	WC-R4 Delta	11.2%	32.1%	40.8%	15.7%	0.2%
	WC-R4 Forest	11.0%	35.0%	35.3%	17.7%	1.1%
Plot method	CP-R2	18.3%	21.3%	48.9%	11.5%	0.0%
	WC-R3	0.8%	7.8%	31.7%	43.6%	16.1%
	WC-R4 Delta	7.8%	36.9%	30.5%	24.8%	0.0%
	WC-R4 Forest	4.9%	34.8%	45.5%	12.8%	2.0%

Transect patches summary and comparison

A comparison of the qualitative transect patch data among reaches indicates that WC-R4 Delta had a more complex and varied lateral profile than the other reaches – especially WC-R3 (figures 9 and 10). WC-R4 Delta and WC-R4 Forest also contained a relatively high amount of detritus (45.3 m and 24.3 m respectively) throughout the reach compared to Camp Polk R2 (11.5 m) and WC-R3 (6 m). WC-R3 had a significantly less complex geomorphic form than the project reaches (figures 8-11) as seen by the relatively narrow area of non-vegetated substrate classes. Indeed, this can also be seen by the number of data points or patch transitions collected in each reach:

CP-R2 N= 69
WC-R3 N= 28
WC-R4 Delta N= 87
WC-R4 Forest N= 64

Higher or lower complexity came in the form of the abundance of wet and dry channels, large log jams, off-channel pools and hydraulic complexity. The relatively simple layout shown for WC-R3 was likely more homogenous than shown here. This is due to the fact that our sample included a dry relic channel within the floodplain that is hydrologically disconnected from the stream except during extreme runoff events. The diversity indices of all three methods may likewise overestimate heterogeneity of this reach.

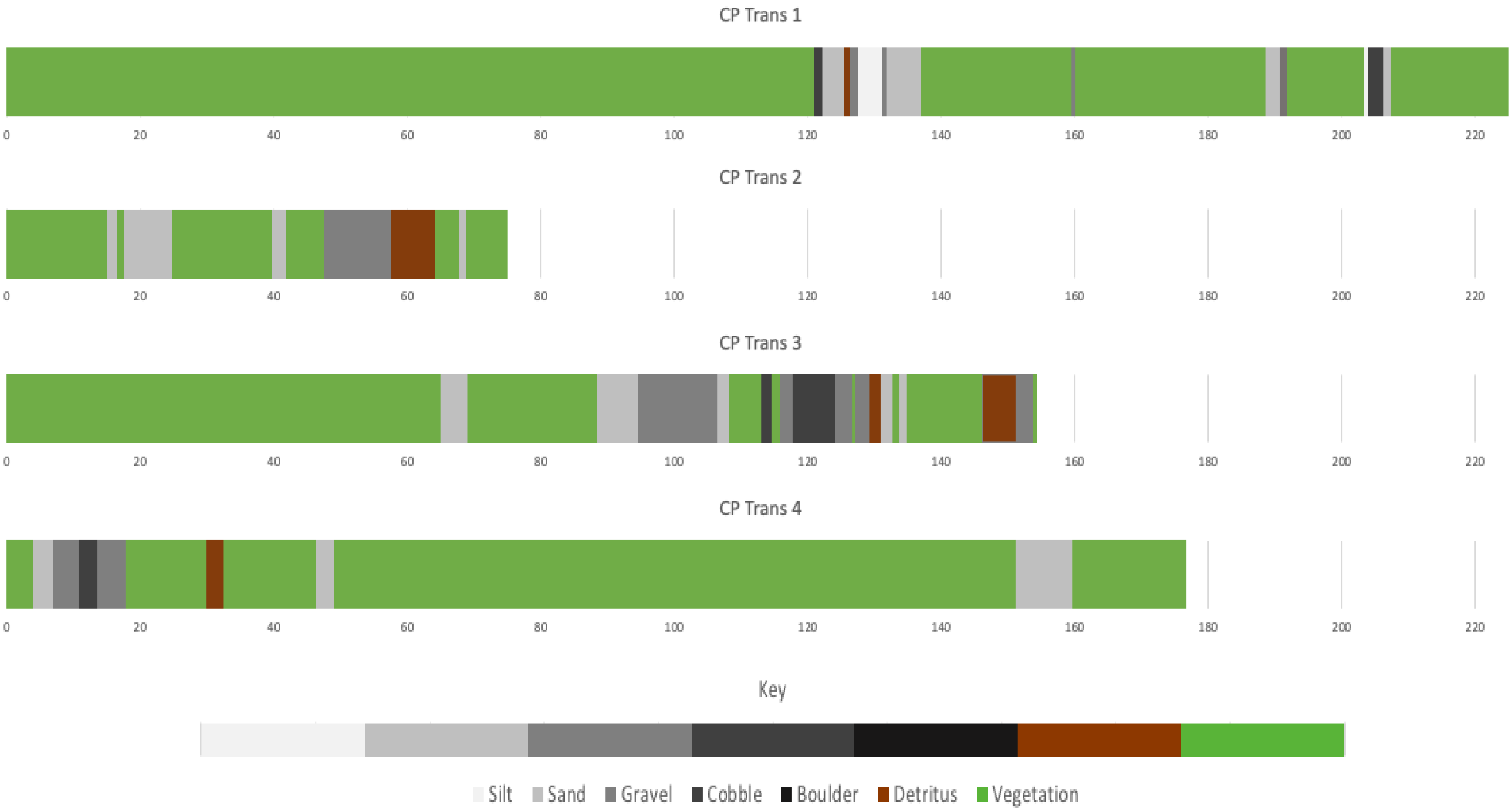


Figure 8: Lateral substrate profile of CP-R2 measured in meters

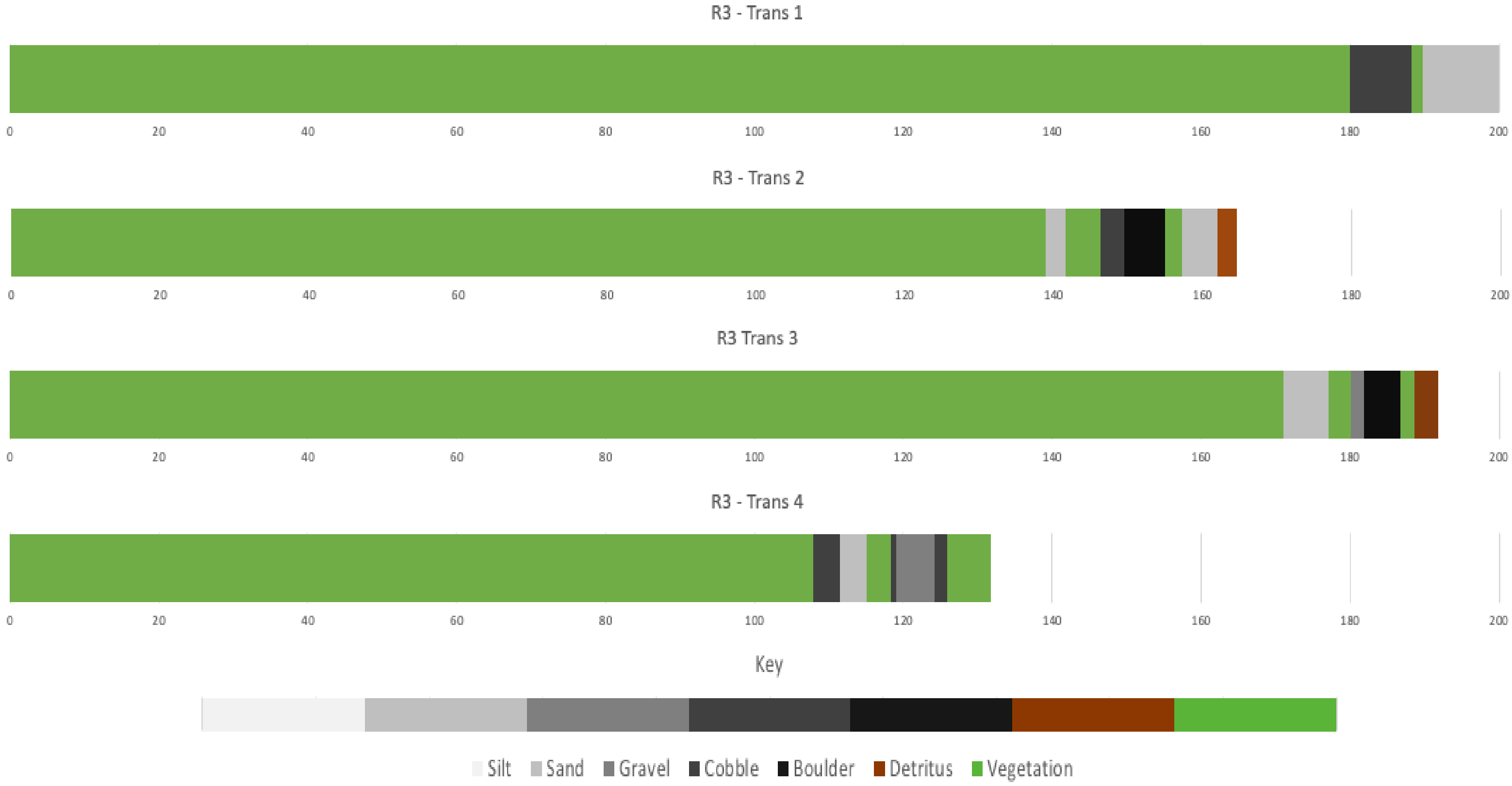


Figure 9: Lateral substrate profile of WC-R3 measured in meters

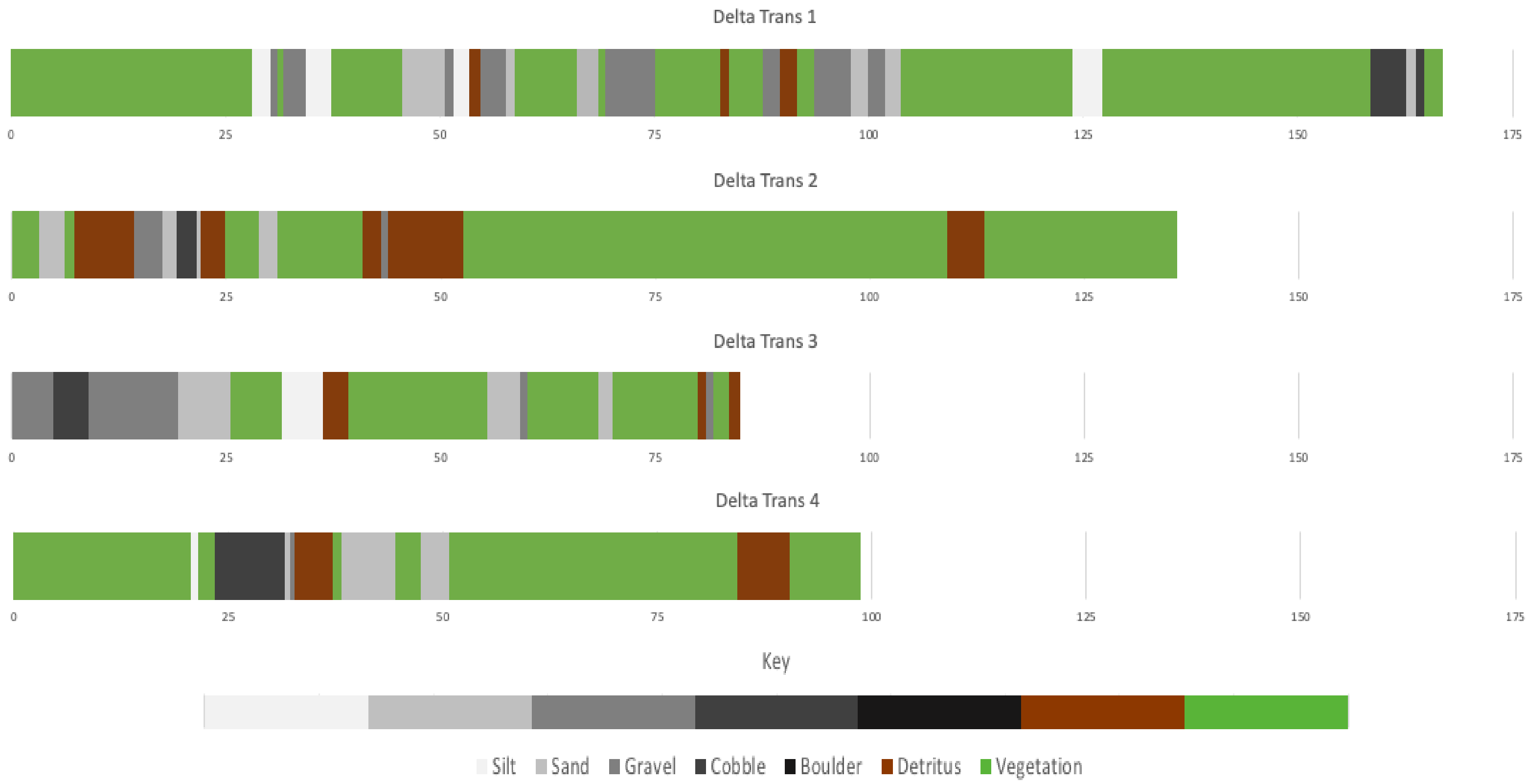


Figure 10: Lateral substrate profile of WC-R4 Delta measured in meters

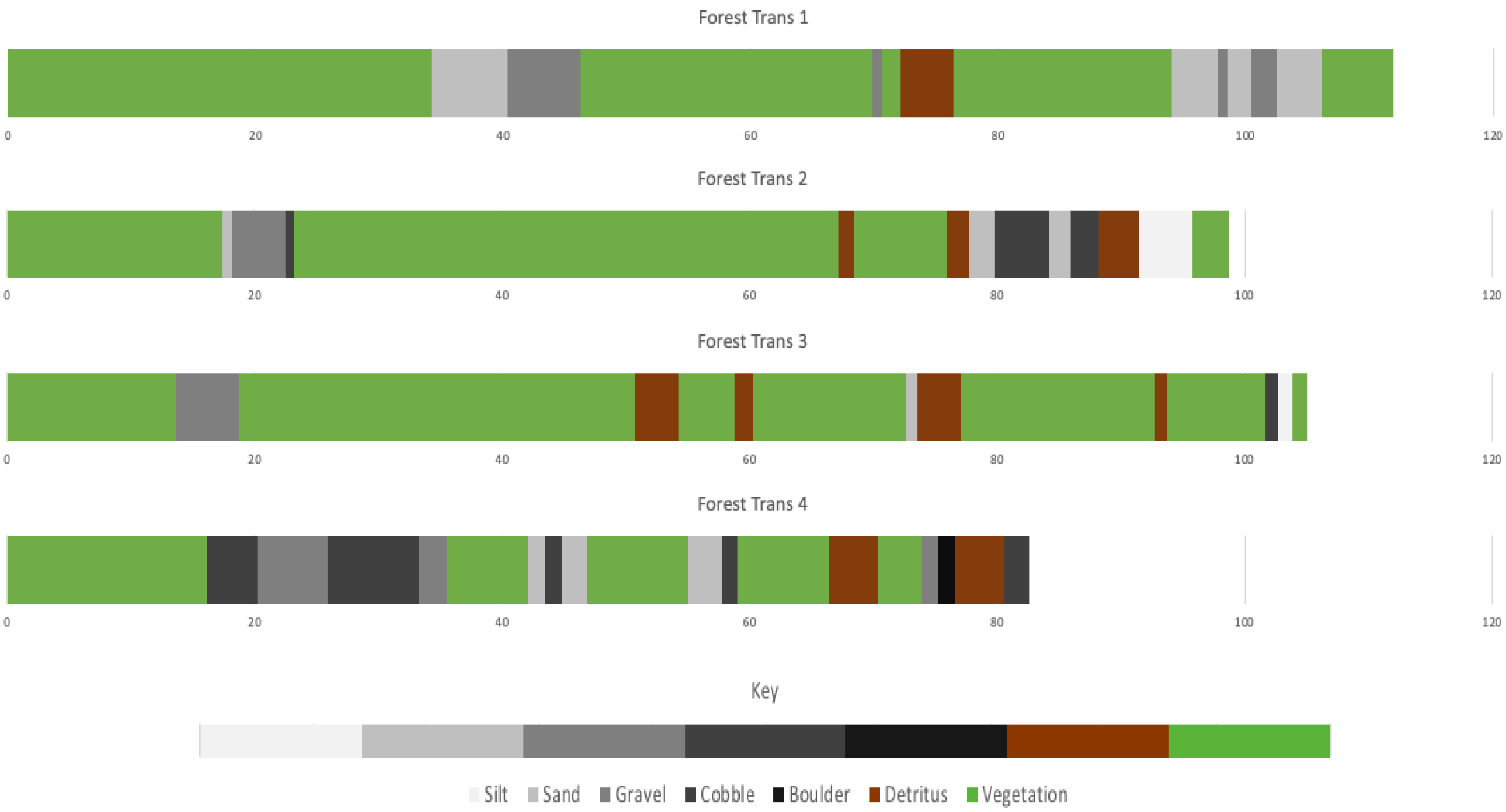


Figure 11: Lateral substrate profile of WC-R4 Forest measured in meters

Statistical testing between reaches

For the pebble count, the chi-squared test indicates that the four reaches were significantly different from each other in terms of substrate class distribution. This test of independence was: X^2 (18, *N* = 1865) = 210.23, *p* <.001. However, for the boulder size class in WC-R3 (5% of the total) we were not able to reject the null hypothesis of independence, indicating this size class in this reach was too similar to the expected frequency and therefore could not be said to be dependent on the reach. This is likely due to the fact that the sample size for WC-R3 was relatively small (N=174) because of the small and confined area of mineral substrates throughout the reach. To reject the null hypothesis in the chi-square test of independence, X^2 must be greater than 5, but for the boulder class in WC-R3, $X^2 = 2.4$. This suggests that larger sample sizes are required in WC-R3 for future years. However, the results for the reach were still significantly different than the other three. The critical value for $X,^2$ *df* =18 at p <.001, is 42.31 so the given value of 210.23 shows strong independence between all reaches.
For the plot method, X^2 (18, *N* = 2306) = 836.57, *p* <.001. In this case, none of the classes in any reach had an expected count <5 which indicates that the sample size was indeed large enough for this method. Once again, the critical value at *p* <.001 is 42.31, therefore the given score of 836.57 indicates strong independence among reaches.

The Cramer's V value for the pebble count method was V=0.19, p < 0.001 which shows that there was a weak - moderate relationship whereas with the plot method, V= 0.35, p <0.001 which indicates that there was a strong to very strong relationship between reach and substrate composition. This means that there is a more significant relationship between substrate compositions and a specific reach under the plot method than the pebble count method.

Quantifying patch heterogeneity - pebble count

Under the pebble count, the heterogeneity metrics indicate that WC-R3 scored lowest on all four index values with an average score of 0.67 (table 6 and figure 12). The indices also illustrate that WC-R4 Forest was the most heterogeneous reach on three of the four outputs with an average score of 0.76. There was a relatively large disparity in scores from the two spatial-specific indices (LIP and Spl.DI) between WC-R3 and the three project reaches with a mean difference of 0.15 (24%). Conversely, there was a relatively small difference between WC-R3 and the others from the two indices *based on reach totals* (SDI and SEI) with a mean difference of 0.05 (8%). However, when the *average group-level* scores of SDI and SEI were compared between the project reaches and WC-R3 (i.e. meso-habitat diversity), the project reaches scored 0.08 higher (12% more diverse). Both the spatial-specific diversities as well as the *group-level* SDI / SEI indices showed greater disparities between the project reaches and the untreated reach than did reach-wide SDI / SEI scores. This indicates that there was more spatial diversity among more substrate classes in the project reaches than WC-R3 (figure 13).

Table 6: Diversity scores between reaches – pebble count

	SDI	SEI	LIP	μ (Spl.DI)
CP-R2	0.69 ± 0.03	0.82 ± 0.05	0.80	0.74
WC-R3	0.66 ± 0.05	0.75 ± 0.1	0.60	0.65
WC-R4 Delta	0.69 ± 0.05	0.79 ± 0.04	0.80	0.73
WC-R4 Forest	0.71 ± 0.02	0.83 ± 0.04	0.80	0.75

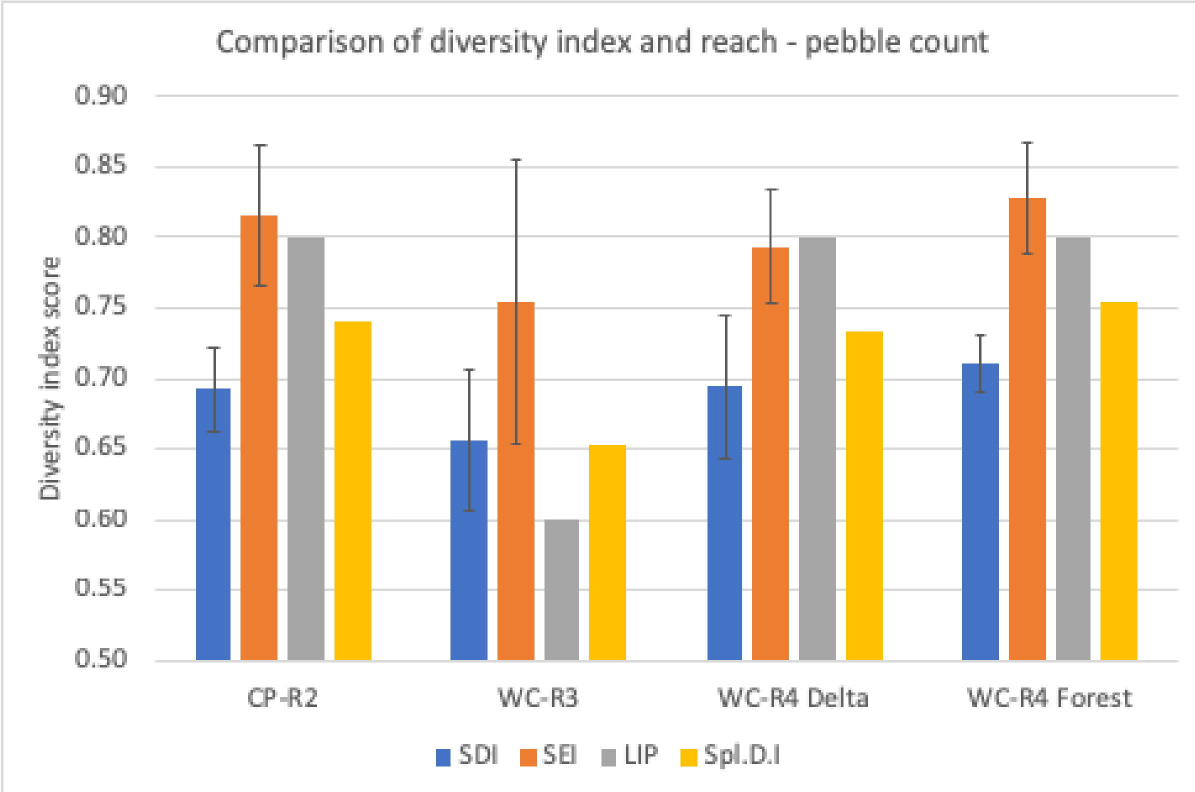


Figure 12: Pebble count diversity indices – all reaches

Table 7: Spatial distribution indices by substrate class - pebble count

	CP-R2		**WC-R3**		**WC-R4 Delta**		**WC-R4 Forest**	
	LIP	Spl.D.I	LIP	Spl.D.I	LIP	Spl.D.I	LIP	Spl.D.I
silt	0.13	0.82	2.67	0.07	0.59	0.70	0.16	0.71
sand	0.23	0.95	0.95	0.73	0.41	1.03	0.13	1.05
gravel	0.25	1.02	0.24	0.99	0.90	1.02	0.07	1.06
cobble	0.18	0.78	0.08	0.97	0.17	0.87	0.61	0.82
boulder	9.56	0.12	1.85	0.50	2.67	0.03	4.80	0.14

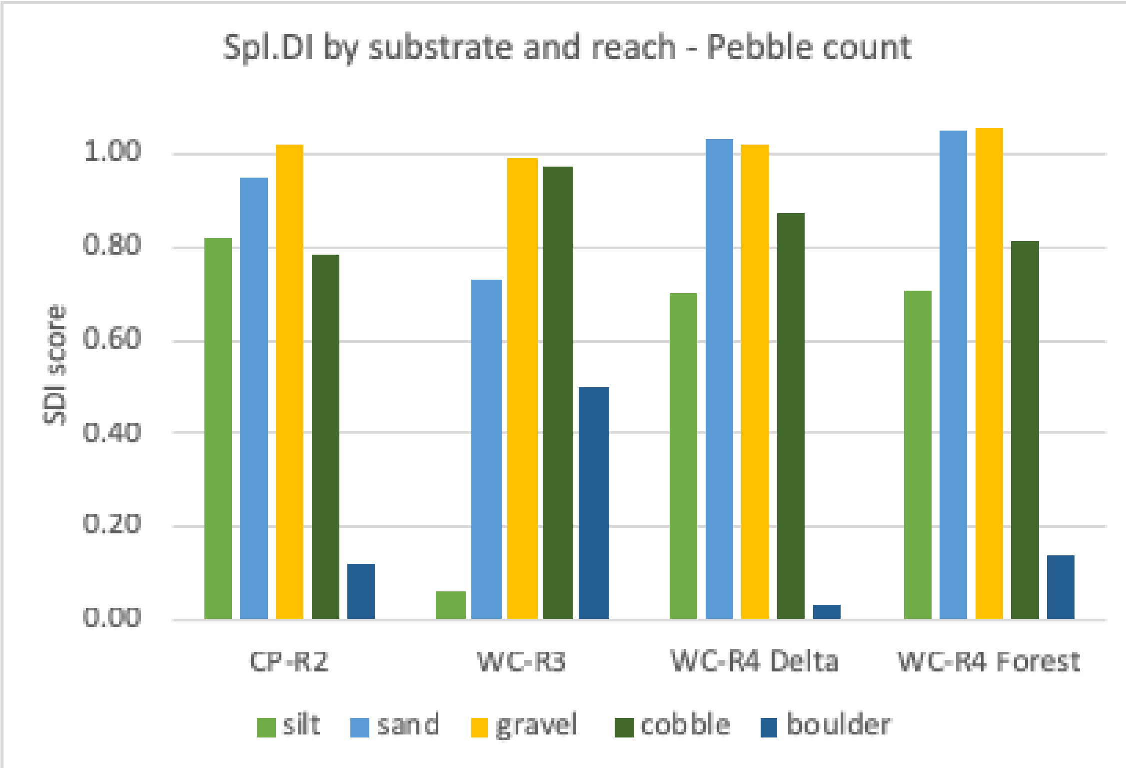


Figure 13: Breakdown of Spatial Diversity Index by substrate class – pebble count

Note that some values under the Spl.DI were over 1.0 which is likely due to rounding errors since the values are combinations of several separate formulas. However, they should be viewed as having a Spl.DI value of 1.0 which indicates a consistent, even dispersal across the entire reach.

Quantifying patch heterogeneity – plot method

Under this sampling method, WC-R4 Delta was on average 5% more diverse than the Forest reach – a reversal of the Delta reach being an average of 2% more diverse than the Forest reach under the pebble count method. The plot method illustrates generally higher scores for the two WC-R4 reaches than the pebble count while simultaneously indicating lower scores in CP-R2 and WC-R3. This method also displays a more distinct substrate signature for each reach than the pebble count. This is illustrated by comparing the wider range of index values in table 8 to the index values in table 6 which show more moderated scores (smaller ranges). This difference in index value range was also supported by the significant differences in Cramer's V test values where under the pebble count, v = 0.19 and under the plot method, v = 0.35 which suggests a more distinct signature for each reach when measured with plots.

Table 8: Diversity scores between reaches – plot method

	SDI	SEI	LIP	μ (Spl. DI)
CP-R2	0.67 ± 0.02	0.66 ± 0.03	0.57	0.54
WC-R3	0.68 ± 0.01	0.65 ± 0.03	0.43	0.51
WC-R4 Delta	0.76 ± 0.01	0.81 ± 0.03	0.86	0.74
WC-R4 Forest	0.74 ± 0.01	0.80 ± 0.04	0.71	0.77

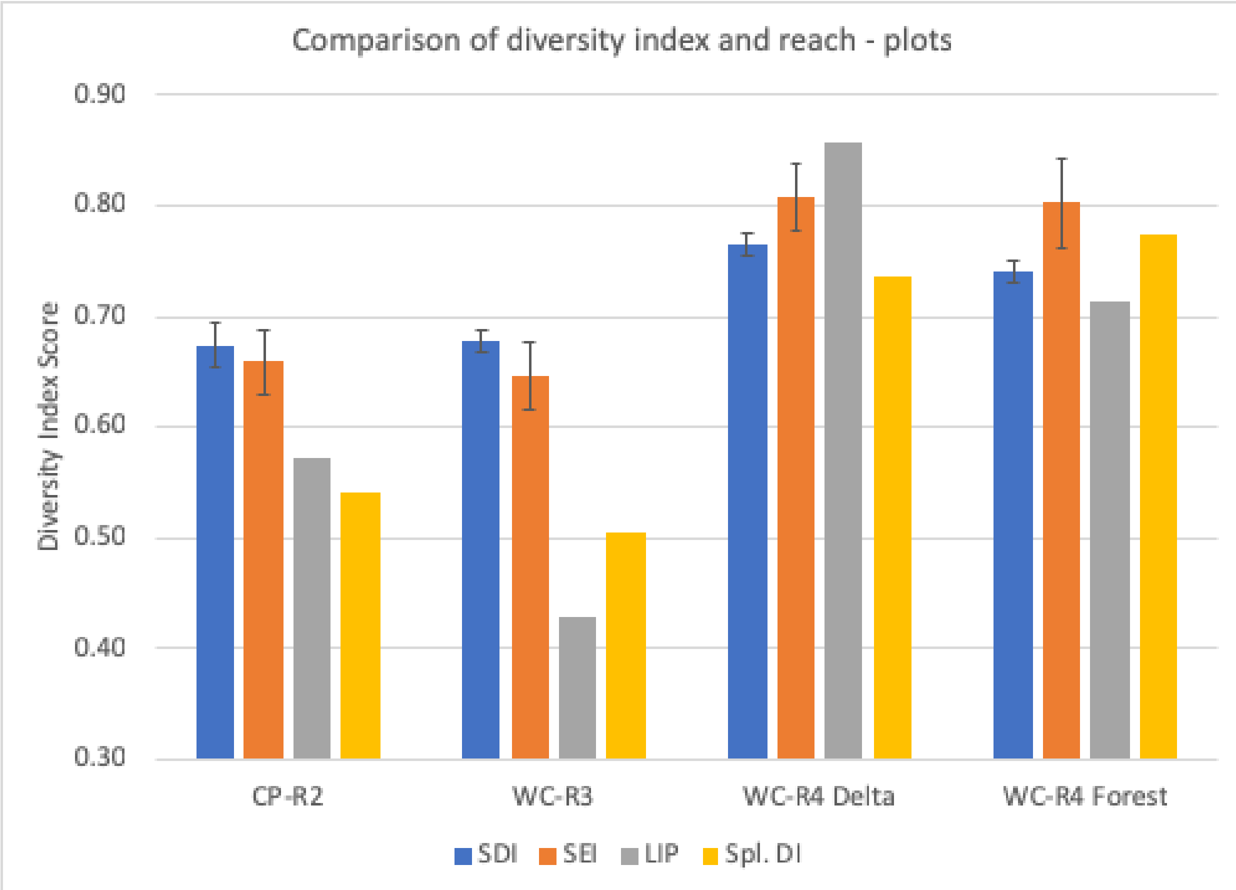


Figure 14: Plot method diversity indices – all reaches

Table 9: Spatial distribution indices by substrate class -plot method

	CP-R2		**WC-R3**		**WC-R4 Delta**		**WC-R4 Forest**	
	LIP	Spl.D.I	LIP	Spl.D.I	LIP	Spl.D.I	LIP	Spl.D.I
silt	0.08	0.98	5.82	0.00	0.50	0.79	1.48	0.50
sand	0.19	0.95	1.01	0.69	0.43	0.87	0.07	0.98
gravel	0.02	1.00	0.27	0.92	0.45	0.85	0.20	0.94
cobble	0.60	0.86	0.09	0.98	0.24	0.92	0.69	0.81
boulder	0.00	0.00	0.24	0.94	0.00	0.00	1.69	0.50
detritus	0.00	0.00	0.00	0.00	0.65	0.82	0.79	0.73
vegetation	5.25	0.00	0.00	0.00	0.43	0.89	0.22	0.96

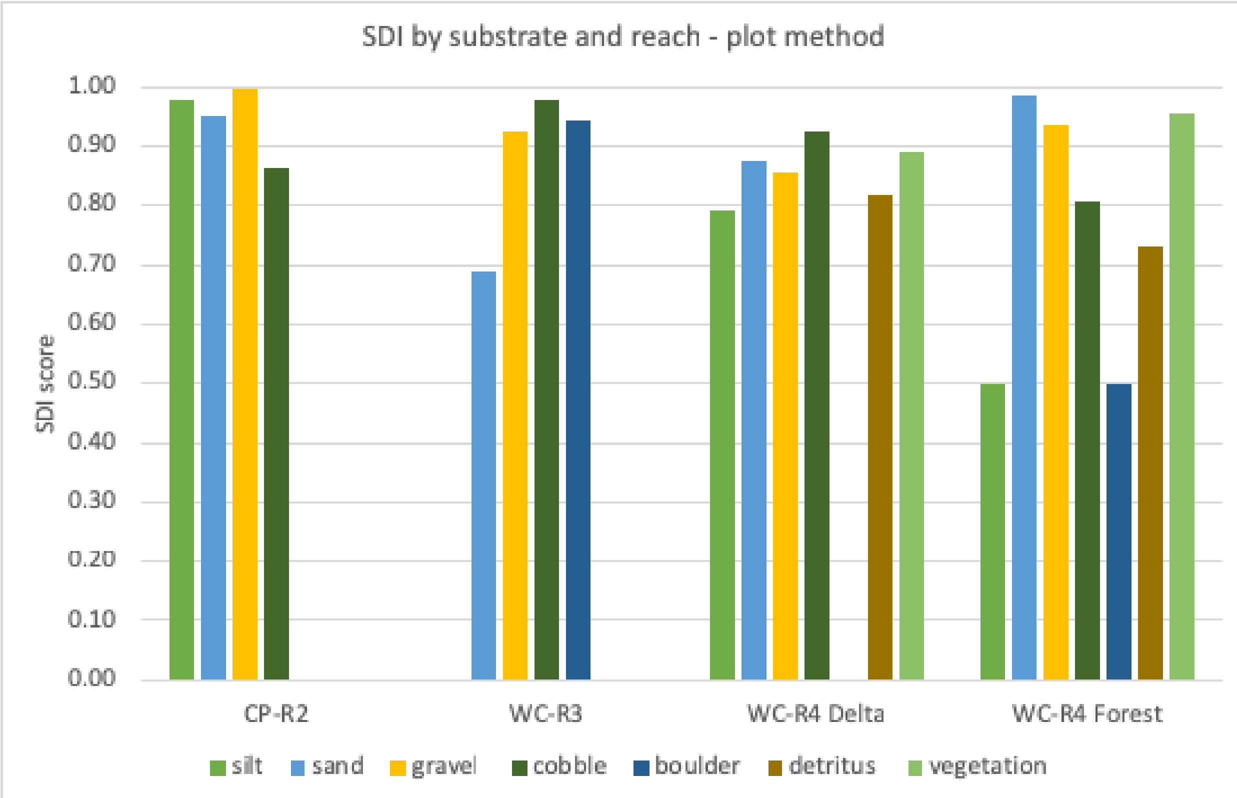


Figure 15: Breakdown of Spatial Diversity Index by substrate class – plot method

Quantifying patch heterogeneity – transect patch method

The transect patch method shows comparatively lower values than the other two methods. This is likely due to the fact that it was the only method that sampled the large areas of vegetation within the active floodplain and the resulting proportions negatively skewed the diversity indices. For example, the CP-R2 transects contained 80.4% vegetation and the other 19.6% of the area was split between six other classes. Nevertheless, it served as an important sampling strategy because the data included the entire transect and did not rely on chance and randomization to represent stream group conditions. In general, the results mirror what the other two methods indicate – that substrata in the two WC-R4 segments were more heterogeneous while WC-R3 remained the most homogenous (table 10 and figure 16).

Table 10: Diversity scores between reaches – transect patch method

	SDI	**SEI**	**LIP**	**Spl.DI**
CP-R2	0.34 ± 0.03	0.36 ± 0.09	0.43	0.62
WC-R3	0.19 ± 0.02	0.22 ±0.08	0.29	0.58
WC-R4 Delta	0.54 ± 0.03	0.60 ± 0.09	0.57	0.71
WC-R4 Forest	0.43 ± 0.03	0.64 ± 0.11	0.43	0.64

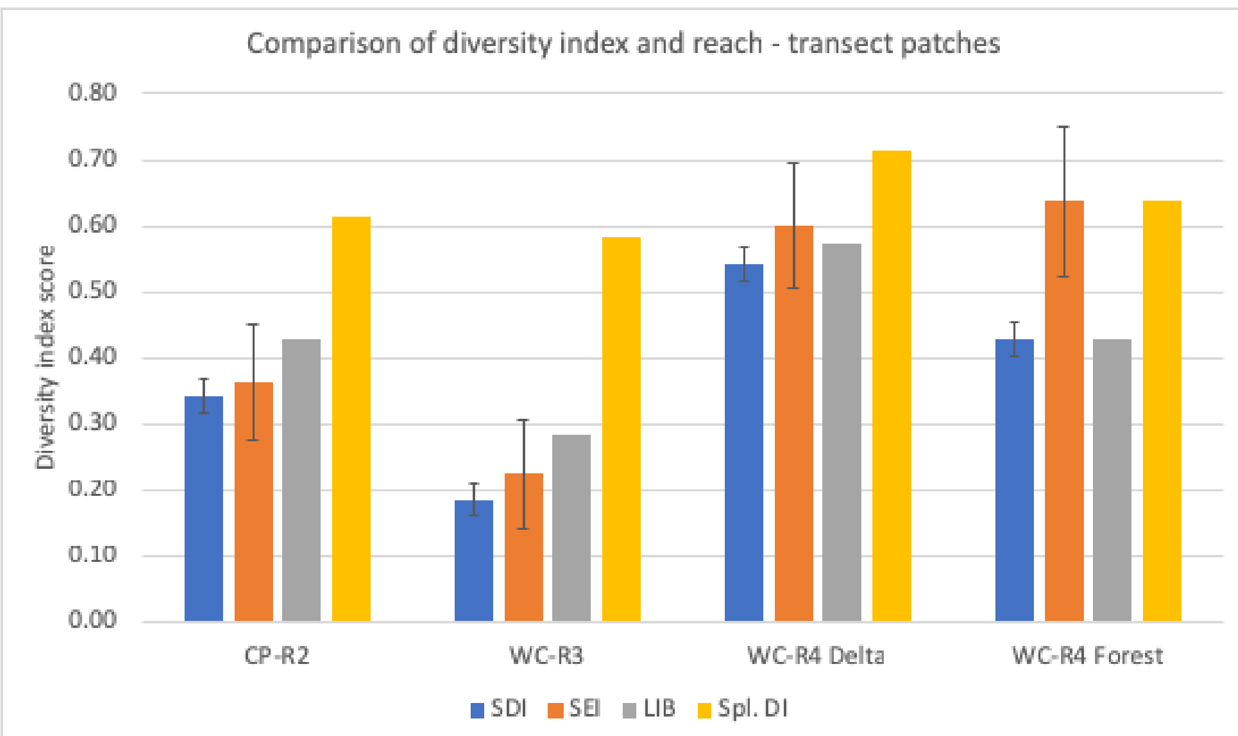


Figure 16: Transect patch method diversity indices – all reaches

<u>Combined methods</u>

Average scores between all methods and indices (table 11) show that WC-R4 Delta was the most diverse reach with a mean score of 0.72 and WC-R3 was the least diverse with an mean score of 0.52 – a 38% difference. Though averaging scores among indices and methods offers an overly simplistic view of each reach, it does provide an overall summary of reach conditions and it allows for a convenient ranking system that includes all types of collected data.

Table 11: Mean diversity scores between methods and reaches. Values for each method are the mean score of SDI, SEI, LIP and Spl.DI for each reach and the mean score among the three methods is again averaged to get a single value for each reach that includes all diversity indices and all methods.

	<u>Pebble Count</u>	<u>Plot</u>	<u>Transect Patches</u>	**<u>Mean score</u>**
CP-R2	0.76	0.61	0.44	**<u>0.60</u>**
WC- R3	0.67	0.56	0.32	**<u>0.52</u>**
WC - R4 Delta	0.76	0.79	0.61	**<u>0.72</u>**
WC - R4 Forest	0.77	0.76	0.52	**<u>0.68</u>**

Correlations between indices and reaches

To test correlations between the data collection method used and diversity indices, we ran a simple Pearson's correlation test. With testing all indices, $r = 0.57$, $N= 16$, $p < 0.05$ which is a moderate correlation. However, since LIP tends to be interpreted on a dichotomous scale (LIP = <1 or >1) to indicate congregation or dispersal, and it is not generally interpreted on a continuum, we ran a separate correlation on only SDI, SEI and μ(Spl. DI). The results show a closer relationship with $r = 0.65$, $N=12$, $p < 0.05$ which is considered a strong relationship (figure 17). Finally, we ran a correlation on all of the mineral Spl.DI scores (biotic classes were dropped from the plot method) for each class and reach – rather than using the mean Spl.DI used in preceding correlations (figure 18). This resulted in the strongest relationship with $r = 0.89$, $N=20$, $p <0.01$. This strong relationship indicates that the two methods were more robust in determining spatial distribution between groups of transects than they were when data was summarized by reach. It also indicates that the two methods captured similar proportions of each substrate class on the group-level. However, as the summary statistics indicate, the pebble count attenuated differences in the between reaches while the two ocular estimate methods accentuated them.

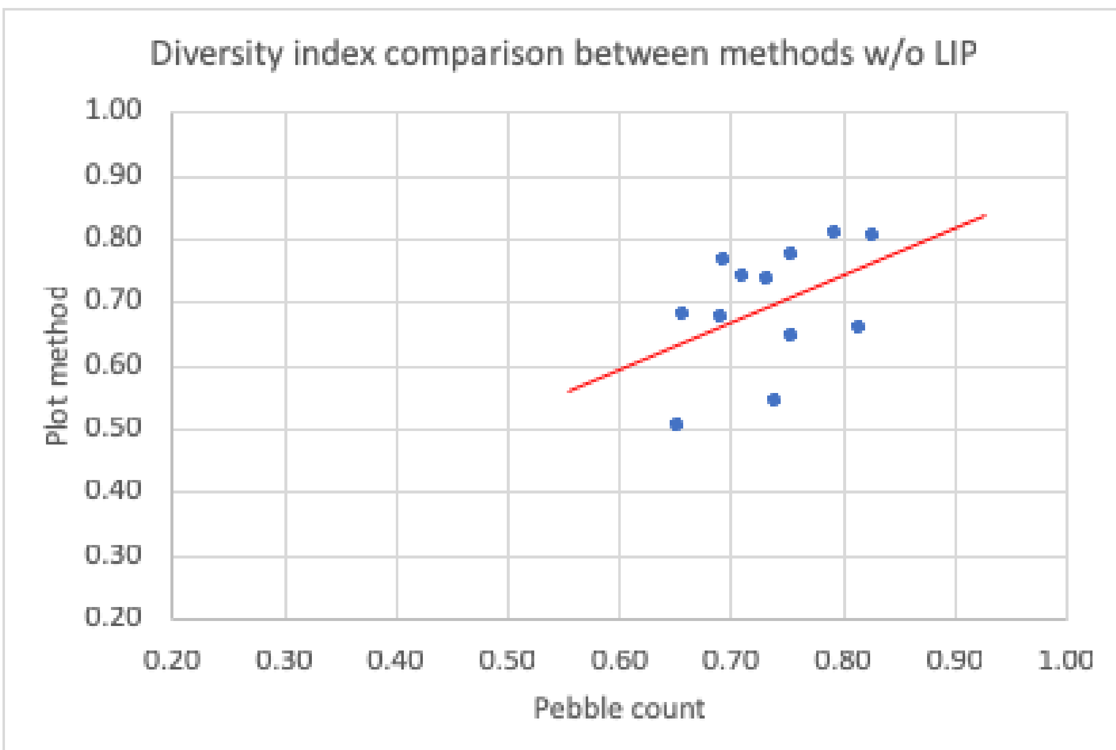


Figure 17: Scatterplot and trendline showing the relationship between pebble count diversity indices and plot method diversity indices without LIP scores. In this example, r = 0.65.

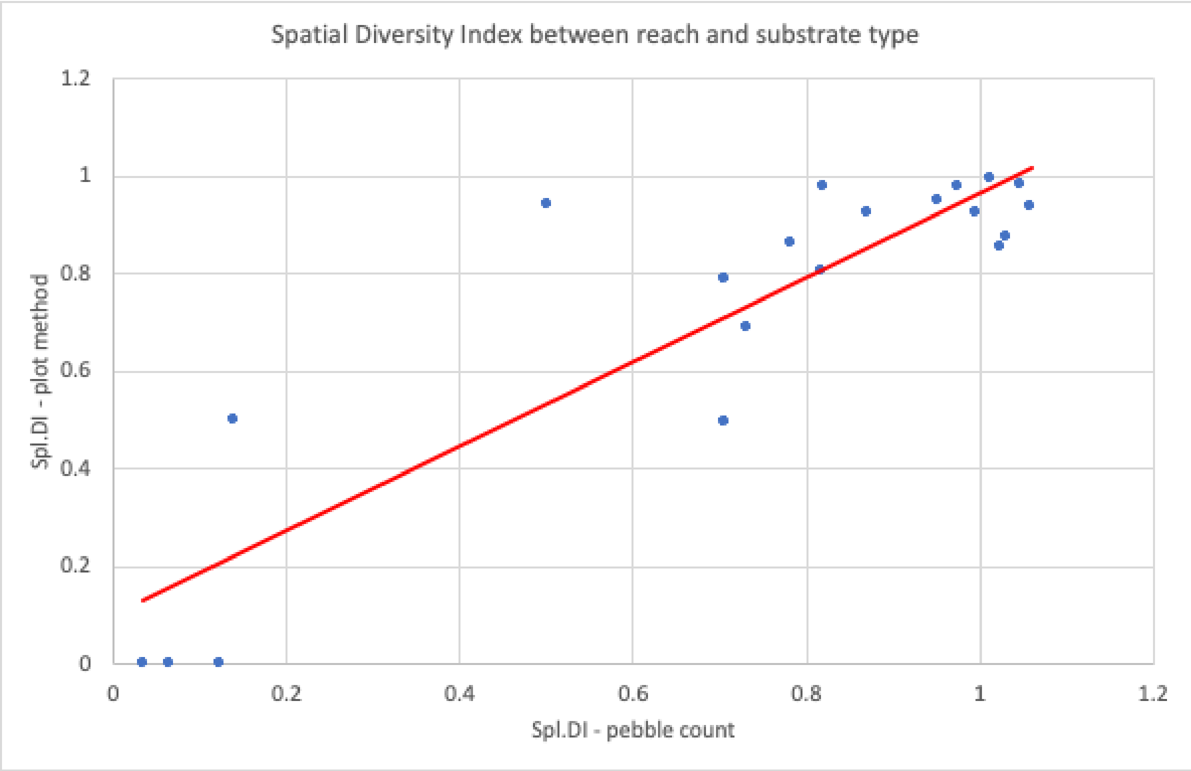


Figure 18: Scatterplot and trendline showing the relationship between pebble count Spl.DI values and plot method Spl.DI values (mineral only). In this example, r = 0.89.

Conclusion

Reach conditions

The methods used for determining substrate patch heterogeneity exhibit clear trends among the four study reaches of Whychus Creek. Quantitative and qualitative data indicate that both WC-R4 sections were the most heterogeneous landscapes. This is somewhat contrary to the original hypothesis that CP-R2 would have the most heterogeneous substrate patches, WC-R4 (both sections together) would have slightly less and WC-R3 would have the most homogeneous distribution of patches. It is encouraging that the most recent restoration project shows the most diverse mosaic of substrata because this should encourage strong rates of post-disturbance recolonization by riparian vegetation, macroinvertebrates, fish and other species at higher trophic levels[22]. However, it is important not to deduce that one reach is necessarily a "healthier" system than another based solely on these results. To illustrate this, WC-R4 had instances of head-cuts, terraced banks, cut banks and other features characteristic of a dynamic stream system progressing through stream evolutionary stages. Conversely, CP-R2 was consistently valued as having a slightly less heterogeneous patch mosaic than WC-R4, even though it has had more time to progress towards a late-evolutionary-stage system. It is therefore important to analyze the findings of this study in the context of other ongoing monitoring efforts throughout Whychus Creek. Whether the geomorphic profile of WC-R4 is and continues to be inherently more diverse or whether its diversity will be attenuated by vegetation recolonization and increased sediment sorting remains to be seen.

There was a clearer distinction between Whychus Canyon reaches 3 and 4. This is especially relevant because the entire canyon resembled WC-R3 prior to the completion of the major restoration activities in 2016. The results presented here clearly illustrate that there was greater substrate heterogeneity in the post-project than pre-project condition. Since WC-R3 can act as a control site, we can infer that the quantitative outputs of this protocol determine how much the restoration activities have created a more diverse landscape. The results of this project can also act as a baseline measurement for a before-after (BA) design of evaluating future restoration activities on WC-R3. However, the data for WC-R3 likely indicated a more diverse habitat than was actually available for aquatic and riparian biota since the transects in this reach extend into a floodplain that is hydrologically disconnected from the creek except during extreme flood events. Specifically, data collected from a relic channel and a discontinued irrigation canal within the boundary of the sample area likely increased the values of heterogeneity.

Data collection and analysis

Based on the results of the chi-square tests, the Cramer's V tests and the diversity indices themselves, it is apparent that the plot method achieved the most accurate quantifiable representation of patch diversity in each reach. The plot method showed stronger relationships between distinct substrate compositions in each reach as well as more significant discrepancies in diversity index scores between reaches. However, the other two methods are useful in other

[22] Milesi, Doledec and Melo. 2016.

contexts. The pebble count allows for a host of additional analytical tools not used in this study[23]. In addition, tracking the gross substrate sizes over time and analyzing size distributions from box-whisker plots can be important indicators of changes in fluvial processes that other two methods will not necessarily capture. The transect patch method produces useful qualitative charts that can indicate how diverse the large patches are in the lateral profiles of the stream. In short, though the plot method seems to exhibit the most accurate representation of patch heterogeneity, all three methods taken together provide a fuller picture of overall conditions that the plot method cannot reproduce in isolation.

Implications

This substrate monitoring protocol has proven to be an informative tool that allowed us to assess the conditions and evolution of fluvial geomorphic processes on Whychus Creek. Stream substrate patchiness and heterogeneity holds important implications for both aquatic and terrestrial biota through niche diversification, ecosystem resilience, post-disturbance recovery and ontogenetic migration[24]. By increasing hydraulic, geomorphic and topographic heterogeneity, the restoration projects in Camp Polk and Whychus Canyon are enhancing these benefits and restoring these reaches to highly productive ecosystems.

Substrate conditions are among the first factors to change within a watershed and being able to track such changes can signal large-scale shifts within the basin. Such changes in land use, water management or natural cycles in the hydrograph are important factors track for the restoration partners. Whereas biotic variables tend to have a longer lag time in responding to such input variables, substrate conditions change on a much shorter timescale, allowing for more immediate assessment and management response. Therefore, by using this pilot project as a baseline, further annual monitoring will be able to track and quantify geomorphic conditions in future years and may help inform restoration-related decision making for other projects.

---

[23] Bunte and Abt 2001.
[24] Milesi, Doledec and Melo 2016.

References

Beisel, Jean-Nicolas, Philippe Usseglio-Polatera and Jean-Claude Moreteau. 2000. "The spatial heterogeneity of a river bottom: a key factor determining macroinvertebrate communities." *Hydrobiologia* 422 (April): 163-171. https://doi.org/10.1023/A:1017094606335

Boyero, Luz. 2003. "The quantification of local substrate heterogeneity in streams and its significance for macroinvertebrate assemblages." *Hydrobiologia* 499, no. 1-3 (June): 161-168.

Brown, Bryan L. 2003. "Spatial heterogeneity reduces temporal variability in stream insect communities." *Ecology Letters* 6, no. 4 (April): 316-325. https://doi.org/10.1046/j.1461-0248.2003.00431.x

Bunte, Kristen and Stephen Bunte. 2001. *Sampling surface and subsurface particle-size distributions in wadable gravel- and cobble-bed streams for analyses in sediment transport, hydraulics, and streambed monitoring.* Fort Collins; U.S. Forest Service. https://www.fs.fed.us/rm/pubs/rmrs_gtr074.pdf

Cluer, Brian and Colin Thorne. 2014. "A stream evolution model integrating habitat and ecosystem benefits." *River Research and Applications* 30, no. 2 (Februrary): 135-154. https://doi.org/ 10.1002/rra.2631

Collins, Scott L. and Melinda D. Smith. 2006. "Scale-dependent interaction of fire and grazing on community heterogeneity in tallgrass prairie." *Ecology* 87, no. 8 (August): 2058-2067. https://doi.org/ 10.1890/0012-9658(2006)87[2058:SIOFAG]2.0.CO2

Dietrich, William E., James W. Kirchner, Hiroshi Ikeda and Fujiko Iseya. 1989. "Sediment supply and the development of the coarse surface layer in gravel-bedded rivers." *Nature* 340: 215. https://doi.org/10.1038/340215a0

Fortin, Marie-Josee, Sergej Payette and Kim Marineau. 1999 "Spatial vegetation diversity index along a postfire successional gradient in the northern boreal forest." *Ecoscience* 6, no. 2 (January): 204-213. https://doi.org/10.1080/11956860.1999.11682521

Li, H. and J.F. Reynolds. 1995. "On definition and quantification of heterogeneity." *Oikos* 73, no. 2 (June): 280-284. https://doi.org/ 10.2307/3545921

Meyer, Kate. 2017. *Deer Creek floodplain restoration project: A Stage 0 Restoration Case Study in Western Oregon*. Report. Portland, OR: U.S. Forest Service, Willamette National Forest. http://www.rrnw.org/wp-content/uploads/3.2FINAL_Deer-Creek-Floodplain-Restoration-Project_RRNW.pdf

Milesi, Syliva V., Sylvain Doledec and Adriano S. Melo. 2016. "Substrate heterogeneity influences the trait composition of stream insect communities: an experimental in situ study." *Freshwater Science* 35, no. 4 (December): 1321-1329. https://doi.org/ 10.1086/688706

Mork, L. and R. Houston. 2016. "2015 Whychus Creek monitoring report." Last modified December 2016. http://www.upperdeschuteswatershedcouncil.org/wp-content/uploads/2018/01/2015-Whychus-Creek-Monitoring-Report_FINAL.pdf

Pederson, Morten and Nikolai Friberg. 2007. "Two lowland stream riffles – linkages between physical habitats and macroinvertebrates across multiple spatial scales." *Aquatic Ecology* 41, no. 3: 475-490. https://doi.org/ 10.1007/s10452-004-1584-x

Statsdirect. 2018. "Diversity of classes." Accessed 15, October 2018. https://www.statsdirect.com/help/nonparametric_methods/diversity.htm

U.S. Forest Service 2012. *Stream inventory handbook: level I & II.* U.S. Forest Service. PDF.

Winemiller, Kirk O., Alexander S. Flecker and David J. Hoeinghaus. 2010. "Patch dynamics and environmental heterogeneity in lotic ecosystems." *Journal of the North American Benthological Society* 29, no. 1 (March): 84-99. https://doi.org/ 10.1899/08-048.1

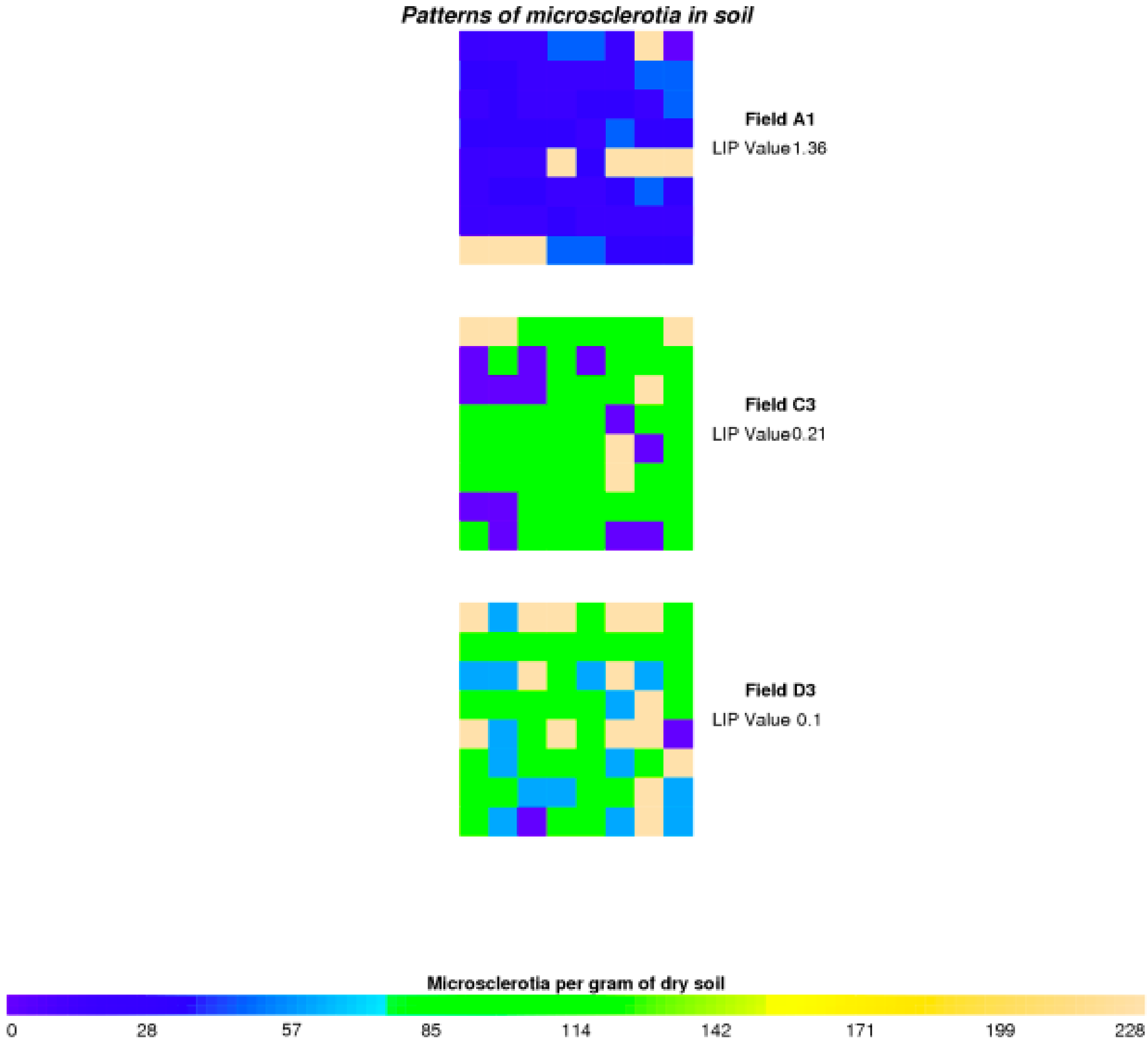
Xiao, C.L., J.J. Hao and K.V. Subbarao. 1997. "Case study #2: Lloyd's Index of Patchiness." Accessed October 13, 2018. http://www.apsnet.org/edcenter/advanced/topics/EcologyAndEpidemiologyInR/SpatialAnalysis/Pages/CaseStudy2Lloyd%27sIndexofPatchiness.aspx

Yabuki, Tetsuo, Yumi Matsumura and Yoko Nakatani. 2009. *Evaluation of pedodiversity and land use diversity in terms of the Shannon Entropy*. Hokkaito: Rakuno Gakuen University. https://arxiv.org/pdf/0905.2821.pdf

Appendix A: Graphic representation of LIP

A sample representation of Lloyd's Index of Patchiness that measures how congregated or separated samples of a particular substrate class are. This example below is the occurrence of a fungus in soil but can be applied to substrate classes (Xiao 1997). Values >1 indicate dispersed, non-congregated dispersal and values <1 indicates a congregated, *patchy* distribution.

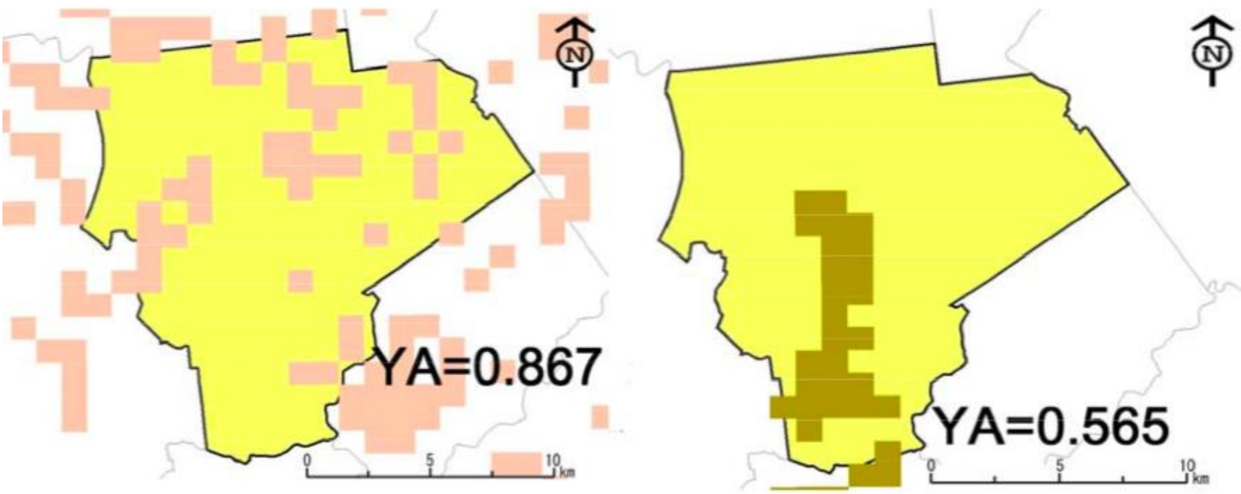
Source: Xiao, Hao and Subbarao 1997[25]



[25] Xiao, C.L., J.J. Hao and K.V. Subbarao. 1997. "Case study #2: Lloyd's Index of Patchiness." Accessed October 13, 2018. http://www.apsnet.org/edcenter/advanced/topics/EcologyAndEpidemiologyInR/SpatialAnalysis/Pages/CaseStudy2Lloyd%27sIndexofPatchiness.aspx

Appendix B: Graphic Representation of Spl.DI

A sample representation of the Spatial Diversity Index which determines how widely dispersed a substrate class is throughout the reach. Values range between 0 and 1 with higher values indicating a more widely dispersed substrate class.



Source: - Yabuki 2009[26]

[26] Yabuki, Tetsuo, Yumi Matsumura and Yoko Nakatani. 2009. *Evaluation of pedodiversity and land use diversity in terms of the Shannon Entropy.* Hokkaito: Rakuno Gakuen University. https://arxiv.org/pdf/0905.2821.pdf

Appendix C: Recommendations

This pilot project proved to be a robust means of determining the geomorphic response to restoration projects on Whychus Creek. I strongly encourage further monitoring based on these methods for future years to be able to analyze how the various reaches are responding to both restoration efforts and natural processes. If this monitoring protocol is able to continue, I have several suggestions to improve the process.

1. If there are sufficient funds and as well as time, conducting all three methods described here would be useful to be able to compare results between years. If there is limited time, the pebble count should be dropped from the protocol. As seen from the these results, it is the least useful in determining patch heterogeneity because the method can muddle the differences in patch distribution by selecting anomalous samples in distinct patches of a particular substrate class. However, a pebble counts conducted every two or three years can be informative since it useful for determining changes in the median and interquartile size distributions throughout the reaches (figure 6). Additionally, it allows for a variety of different analyses that detect changing stream conditions.

2. I recommend an increased sample size for the plot method in future years. Rather than the three samples per transect (36 per reach) collected in the pilot project, I propose using six per transect (72 per reach) to more accurately capture the conditions in each group. The transect patch method charts indicate higher diversity of substrate classes and patchiness than reflected in the diversity indices under the plot method so a higher sample size per transect may be able to narrow that difference. Specifically, I recommend using the same randomization process as with the original three-placement strategy but instead each placement would sample two contiguous plots. To reduce bias, the surveyor would sample and record one plot placement, then consistently flip the plot along the transect always to river right or always river left to sample again. Under this design, each of the three placements would sample a 2 x 0.5-m area with 32, 12.5 x 25-cm quadrants counted.

3. If time is very limited to and only one method can be conducted, the plot method is the most useful for quantifying patch heterogeneity, accurately determining significant differences between reaches and efficiently sampling each transect. Two surveyors can gather the 36 samples from each reach in approximately three hours per reach – one hour for WC-R3 in its current state.

4. I recommend that future surveyors read the protocol carefully and cover this report before collecting official data. There are multiple strategies that the surveyor can implement to improve accuracy and replicability – included in the separate protocol document. I also highly suggest that only one surveyor conducts each method, as adding more surveyors increases bias discrepancies. Finally, each surveyor should practice each method along an entire transect before officially gathering data, as this has been shown to improve accuracy.

5. I also recommend that future data analysis include more detailed statistical tools than used here. There are more options available to draw stronger conclusions when there are multiple samples taken from the same sites. Since pre-formatted excel sheets and analytical instructions are included with this pilot project, analyzing the results at the level presented here should not require a large time commitment, and more work can be done to expand upon the groundwork laid out in this report. This is especially true if the UDWC is going to do a BA design analysis of future restoration on WC-R3 because it would be important to understand how consistent or inconsistent the monitoring protocol is between years before the baseline analysis is compared to the post-restoration response.

6. If another similar pebble count is conducted, the surveyor should collect more samples in WC-R3 while it is in its current condition. The chi-square test of independence suggested that the sample size was slightly too small for that reach. If using the sampling frame, each frame placement should sample four pebbles rather than two as was done in 2018. This can easily be done by using the point intersect method of the four outermost crosshairs on the frame.

7. The findings of this project could be compared within a larger context of other sampling protocols such as macroinvertebrate abundances and diversities. It would be very interesting to run regression analyses between patch diversity and macroinvertebrate diversity to see if and how they respond to patch heterogeneity.

8. It will be informative to analyze the meso-habitat heterogeneity of each of the four transect groups in each reach. It would be interesting to determine and then average the meso-habitat diversity values from SDI and SEI. The combined group score will likely be different than the reach score because it is not pooling all the data to determine a single value. However, Spl.DI and LIP should not be calculated on a group level because the data input would have to be calculated from changes among transect-level data of which the sample sizes are likely too small.