2011

Exploration of Stream Habitat Spatial Modeling; Using Geographically Weighted Regression, Ordinary Least Squares Regression, and Natural Neighbor Interpolation to Model Depth, Flow, and Benthic Substrate in Streams

Kenneth R. Sheehan
West Virginia University

Follow this and additional works at: <https://researchrepository.wvu.edu/etd>

Recommended Citation

Sheehan, Kenneth R., "Exploration of Stream Habitat Spatial Modeling; Using Geographically Weighted Regression, Ordinary Least Squares Regression, and Natural Neighbor Interpolation to Model Depth, Flow, and Benthic Substrate in Streams" (2011). *Graduate Theses, Dissertations, and Problem Reports*. 3482.
<https://researchrepository.wvu.edu/etd/3482>

This Dissertation is protected by copyright and/or related rights. It has been brought to you by the The Research Repository @ WVU with permission from the rights-holder(s). You are free to use this Dissertation in any way that is permitted by the copyright and related rights legislation that applies to your use. For other uses you must obtain permission from the rights-holder(s) directly, unless additional rights are indicated by a Creative Commons license in the record and/ or on the work itself. This Dissertation has been accepted for inclusion in WVU Graduate Theses, Dissertations, and Problem Reports collection by an authorized administrator of The Research Repository @ WVU. For more information, please contact researchrepository@mail.wvu.edu.

Exploration of Stream Habitat Spatial Modeling; Using Geographically Weighted Regression, Ordinary Least Squares Regression, and Natural Neighbor Interpolation to Model Depth, Flow, and Benthic Substrate in Streams

Kenneth R. Sheehan

Dissertation submitted to the
Davis College of Agriculture, Natural Resources and Design
at West Virginia University
in partial fulfillment of the requirements
for the degree of

Doctor of Philosophy
In
Forest Resources Science

Approved by
Stuart A. Welsh, Ph.D., Committee Chairperson
Kyle J. Hartman, Ph.D.
Steve Kite, Ph.D.
Walter Kordek, M.S.
Michael P. Strager, Ph.D.

Division of Forestry and Natural Resources
Morgantown, West Virginia

2011

Keywords: Benthic substrate; geographically weighted regression; ordinary least squares; interpolation; spatial scale; modeling; assessment; GIS; stream habitat

Copyright Kenneth R. Sheehan 2011

ABSTRACT

Exploration of Stream Habitat Spatial Modeling; Using Geographically Weighted Regression, Ordinary Least Squares Regression, and Natural Neighbor Interpolation to Model Depth, Flow, and Benthic Substrate in Streams

Kenneth R. Sheehan

Assessment and modeling of stream habitat are integral to understanding streams and the biota within them. In the past several decades, assessment sophistication of ecologic systems increased due to analysis power afforded by gains in computing capability. Spatial data analysis methodology grew alongside computing power and incorporated spatial qualities of ecological data, thereby providing new insights. New methods like geographically weighted regression (GWR) and more established methods like interpolation are now being used in ecological studies to guide assessments and management decisions. However, their accuracy and utility for analysis of stream habitat data have not been fully explored. To clarify their impacts on stream habitat data, the five chapters of this dissertation examined spatial qualities (e.g. heterogeneity, scale, sample pattern) and the use of interpolation and GWR on depth, flow velocity, and benthic substrate.

Benthic substrate, depth, and flow velocity data were collected from four streams between July 2005 and August 2010. Data were collected from Aarons Creek, Monongalia County, WV, Elk River, Kanawha County, WV, Little Wapiti and Grayling creeks in Gallatin County, MT. Using GIS, the datasets were mapped, modeled, and analyzed between fall 2009 and summer 2011.

Results from our studies demonstrated GWR outperformed non-spatial ordinary least squares regression (OLS) when modeling benthic substrate. Our study showed stream data collected at a single scale may be used to generate meaningful results at scales other than that at which it was collected. This finding is important for stream habitat studies where data are often collected at varying spatial scales. As spatial heterogeneity of benthic substrate increased, accuracy levels of models decreased showing heterogeneity must be quantified in analysis of stream habitat variables. Large (>20m width) and small (<10m width) wadeable streams may be analyzed using the same type of spatial analysis though substrate deposition pattern may vary in different size streams. Benthic substrate depositional pattern was most effectively captured by non-random point selection which created more accurate maps than grid and random point sample methods.

Combined results demonstrated the need to address spatial qualities of stream habitat data in analysis, assessment, and how spatial attributes may guide data collection. Further, failure to quantify spatial attributes in stream habitat data can cause erroneous results and thus minimize effectiveness for useful ecologic conclusions and management decisions.

Dedication

I would like to dedicate my dissertation to my parents Richard and Patricia Sheehan, my sister Nancy, and Auntie E. Without their friendship and love, unfailing emotional, financial, and intellectual support I would never have traveled this far in life.

Acknowledgments

Many thanks must be given to those who were integral in helping me finish my degree. I would like to call special notice to those who've inspired me along the way including friends who helped make science and hard work a lot more fun; Chris Horn, Garrett Staines, Ryan Utz, Jason Stolarski, Andy Hafs, Nathan and Spike Ten Eyck, and Jesse De La Cruz. Particular thanks to my Dad for 18 hour days collecting data as my field tech and not getting too upset when I told *him* what to do. Thanks to Nicole Ten Eyck for many hours in quality control and data entry. Thanks to Pat Mazik for keeping me on track; Becky Nestor for being the organizational wizard and calm in the paper-trail storm; Donna Hartman for being a good sounding board; and Katrina Klugh Stewart for allowing me to hone my teaching skills. Thank you to my committee members; Kyle Hartman for providing his grad students who helped me find friendship and sense of place at WVU; Walt Kordek for keeping me on my toes; Mike Strager for introducing me to GIS and giving me opportunities to expand my knowledge and skills; and Steve Kite for helping me mix science with social conscience. Special thanks (of course) go to Stuart Welsh for his foresight and support in allowing me to pursue my own ideas and interests for my degree.

Table of Contents

Chapter 1- Introduction to the Dissertation.....	1
Literature Cited.....	9
Chapter 2- <i>An Effective Data Selection Pattern for Modeling Benthic Substrate in Streams; Using Ordinary Least Squares Residuals to Guide Data Collection</i>.....	22
Abstract.....	22
Introduction.....	23
Methods.....	26
Results.....	29
Discussion.....	31
Acknowledgments.....	36
Literature Cited.....	36
Chapter 3 – <i>Accuracy of Stream Habitat Interpolations Across Spatial Scales</i>.....	47
Abstract.....	47
Introduction.....	48
Methods.....	50
Results.....	54
Discussion.....	57
Acknowledgments.....	62
Literature Cited.....	62
Chapter 4 - <i>Advantages of Geographically Weighted Regression for Modeling Benthic Substrate in Two Greater Yellowstone Ecosystem Streams</i>.....	77
Abstract.....	77
Introduction.....	78
Methods.....	80
Results.....	83
Discussion.....	85
Literature Cited.....	88
Chapter 5- <i>Geographically Weighted Regression Models and Role of Moran's I in Evaluating Heterogeneous Benthic Substrate in Streams</i>	100
Abstract.....	100
Introduction.....	101
Methods.....	103
Results.....	105
Discussion.....	106
Acknowledgments.....	109
Literature Cited.....	110
Chapter 6- <i>Use of Geographically Weighted Regression to Model Benthic Substrate; A Comparison Between Large and Small Wadeable Streams</i>.....	114
Abstract.....	114
Introduction.....	115

Methods.....	119
Results.....	123
Discussion.....	124
Literature Cited.....	134
Chapter 7- <i>Implications of the Research</i>	139
Appendix-<i>Study Site Graphics</i>	144
Study sites located on map of contiguous United States.....	144
West Virginia study sites with aerial overlay.....	145
Montana study sites with aerial overlay.....	146
Curriculum Vitae (CV)	147
References	153

*Note- Some formatting differences, such as citation format, exist between chapters due to specific requirements for journal submission.

List of Tables

Chapter 2

Table 1. Root mean square error values for each sampling method for all sites. Values in bold are closest to the actual coverage for each site..... 38

Table 2. Sampling information and percent of each substrate type predicted successfully at each site by sampling method. Percentages represent exact spatial (x,y coordinate) match. Values in bold indicate the best performing model. Non-random sampling guided by OLS residuals and substrate deposition pattern was the best performing sample pattern.....39

Chapter 3

Table 1. Raster cell size in meters squared for each predictive scale.....67

Table 2. Wapiti Creek base scale (0.3 m) compared to scales 1, 3, 5, 7, 9, and 11. Base scale values of substrate, 1 (silt), 2 (sand), 3 (gravel), 4 (cobble), 5 (boulder), and 6 (land) are plotted on the x axis and comparative scales 1, 3, 5, 7, 9, and 11 are plotted on the y axis. Plots were fit with a regression line and associated r^2 values to show proportional accuracy of scale predictions to the base scale.....68

Chapter 4

Table 1. Ordinary least squares diagnostics indicating the model variables are neither redundant nor exhibit multicollinearity (VIF), are statistically significant (Joint Wald), exhibit non-stationarity (Koenker BP), and exhibit residuals distributed non-normally (Jarque-bera).....92

Table 2. Ordinary least squares regression and GWR r^2 and adjusted r^2 values for depth and flow velocity relationships to substrate on Little Wapiti and Grayling creeks, Montana. Akaike's information criterion values are also included. Depth and flow velocity are explanatory variables, and substrate is the predicted value for all regressions.....93

Table 3. Akaike information criterion parameters from GWR and OLS regressions for Little Wapiti and Grayling creek sites. Explanatory variables were depth and flow velocity. Substrate is the predicted value for all regressions.....94

Chapter 5

Table 1. Raw geographically weighted regression r^2 values in descending order and labeled by site. Clendenin maintained both the highest and lowest r^2 values. Range of Moran's I and r^2 was well distributed across all sites. Moran's I index is of benthic substrate pattern.....112

List of Figures

Chapter 2

Figure 1. Substrate deposition patterns (a) and ordinary least squares regression patterns of over and under estimation of substrate size (b) for Elk, Little Wapiti, Grayling, and Aarons sites. The boundary of over and under estimation closely follows the edge boundaries of substrate transitions zones.....41

Figure 2. Example on the Little Wapiti site of the four point selection types used for interpolations.....42

Figure 3. Natural neighbor interpolations of Little Wapiti created by grid, random, and actively selected points. Maps illustrate the ability of ordinary least squares regression to identify substrate transition zones and aid in the point selection process allowing the method to out-perform grid and random methods using the same amount of data.....43

Figure 4. Natural neighbor interpolations of Grayling Creek data created by grid, random, and actively selected points.....44

Figure 5. Natural interpolations of Aarons Creek data created by grid, random, and non random points. Aarons Creek maintained the most equal distribution of substrate of all sites.....45

Figure 6. Natural neighbor interpolations of Elk River data created by grid, random, and non-random data sampling. Elk River was the least heterogeneous site and was made up of more than 50% sand.....46

Chapter 3

Figure 1. Example of x,y coordinate point grid for Grayling Creek showing 4,950 data locations each containing depth, flow, and dominant observed substrate information. Natural neighbor interpolation of the points created the baseline visual map.....69

Figure 2. Root mean square error values for natural neighbor predicted maps of depth values on Grayling and Wapiti creek. Scales 1-12 are found on table 1 and range from 0.1 to 13.38 m² cell size.....70

Figure 3. Root mean square error values for natural neighbor predicted maps of flow velocity values on Wapiti Creek. Scales 1-12 are found on table 1 and range from 0.1 at the base scale to 13.38 m² cell size at scale 11.....71

Figure 4. Root mean square error values for predicted maps of all substrate values at Wapiti and Grayling creeks. Scales 1-12 are found on table 1 and range from 0.1 to 13.38 m² cell size.....72

Figure 5. Wapiti Creek substrate RMSE values with a logarithmic trend line applied showing r² values for each. Increase in RMSE as scale moves away from the baseline reference scale shows

a progressive tapering effect of accuracy loss as predictive scale moves away from the baseline. Sand is the substrate size range with the smallest RMSE change between scales.....73

Figure 6. Grayling Creek substrate RMSE values with a logarithmic trend line applied showing r^2 values for each. Increase in RMSE as scale moves away from the baseline reference scale shows a progressive tapering effect of accuracy loss as predictive scale moves away from the baseline. Sand and gravel maintained the smallest RMSE changes between scales.....74

Figure 7. Wapiti Creek residual maps showing depth, flow, and substrate standard deviations for each of the 3,630 x,y coordinate points at the site. Highly localized regions of standard deviation variation in scale one progress to larger regions of similar standard deviation values as scale becomes smaller, referred to in the text as a smoothing effect.....75

Figure 8. Grayling Creek residual maps showing depth, flow, and substrate standard deviations for each of the 3,630 x,y coordinate points at the site. Abrupt localized changes in standard deviation are more prevalent in scale one. A more gradual change in standard deviation, or smoothing effect, may be seen in the coarsest scale.....76

Chapter 4

Figure 1. Grayling Creek wetted width at upstream boundary is 17.7m, and 20.1m at the downstream boundary; maximum width is 20.1m and minimum width is 13.9m. Little Wapiti Creek wetted width at upstream boundary is 3.7m, and is 2.3m at the downstream boundary; maximum width is 9.4m and minimum width is 1.8m.....96

Figure 2. Grayling Creek standardized residual maps and predicted benthic substrate from GWR and OLS regressions.....97

Figure 3. Standardized residuals and predicted benthic substrate maps from GWR and OLS regressions.....98

Figure 4. Condition number and local r^2 maps of Little Wapiti and Grayling creeks created by GWR.....99

Chapter 5

Figure 1. The linear trend of Moran's I values decrease as geographically weighted regression r^2 values decrease for 40 stream sub-sites. This relationship shows regression values of substrate are correlated to Moran's I and thus spatial distribution type.....113

Chapter 6

Figure 1. Little Wapiti and Grayling creek locations with 30 random sample points applied within each site. Actual substrate and depth information is also included for reference. Geographically weighted regressions were performed in a 2.5 meter radius surrounding each random sampling point.....130

Figure 2. Aarons Creek and Elk River locations with 30 random sample points applied within each site. Actual substrate and depth information is also included for reference. Geographically weighted regressions were performed in a 2.5 meter radius surrounding each random sampling point. Elk River displays greater areas of substrate homogeneity than other sites.....131

Figure 3. Residuals of OLS (top of figure) and GWR residuals (bottom of figure) of Little Wapiti and Grayling creeks. Visual inspection of standard deviations of GWR contain a much larger amount of results within 0.5 standard deviations, the result of non-static variance applied at individual locations.....132

Figure 4. Residuals of OLS regression (top of figure) and GWR residuals (bottom of figure) of Aarons Creek and Elk River. Elk River has large areas of more homogeneous substrate, which GWR was better able to model using local variance calculations.....133

Appendix

Figure 1. Little Wapiti, Grayling, Elk and Aarons study sites located on a map of the contiguous United States with Montana and West Virginia insets.....144

Figure 2. State of West Virginia with Elk and Aarons site locations with aerial insets of the location of the sites.....145

Figure 3.State of Montana with Grayling and Little Wapiti site locations and aerial insets of the location of the sites.....148

Chapter 1- Introduction to the Dissertation

This dissertation examines the assessment of stream habitat variables (depth, flow and benthic substrate), focusing on the importance of spatial qualities on assessment, modeling and analysis. It further explores scale, spatial heterogeneity, spatial versus non-spatial regression, and patterns of data sampling. In order to understand stream habitat assessments, we must review how habitat variables are valuable to the ecological study of streams.

Stream habitat assessment and modeling have existed in varying forms since the mid 1920's and evolved along with advances in stream science and ecological understanding (White 1996). Assessment and modeling of stream habitat are important because success of a fish population is directly tied to its associated habitat. Assessment and modeling of stream habitat allows fisheries biologists to quantify the type of habitat available at a given time and make inferences about fish communities and aspects of fish-habitat interactions within streams (Ian 1999; Jungwirth et al. 1995; Muhar and Jungwirth 1998; Schlosser 1990). The importance of habitat to fishes cannot be underestimated, as indicated by the building efforts in the last several decades to identify suitable and required fish habitat (Cooper et al. 1998; Guido et al. 2006; Mac Nally and Quinn 1998; Orth and White 1999; Rosenfeld 2003; Scott et al. 2002a; Scott et al. 2002b; Thompson et al. 2001; Townsend et al. 2003). Stream habitat studies have expanded to encompass varying ecological scales within the past half century to aid in understanding fish behavior, occurrence, and habitat use (Karr et al. 1986; Townsend et al. 2003; Winemiller 2005; Winemiller et al. 2010). Not surprisingly, criticism during the 1980's of fish and stream management plans without extensive habitat study and assessment led to more ecosystem wide stream management and focus on habitat quantification (Frissell et al. 1997; Lewis et al. 1996; Orth and White 1999).

Habitat information coordinated with knowledge of fish species facilitates best management decisions (Ian 1999). Further, habitat information is commonly used for stream classification (Montgomery and Buffington 1997; Montgomery et al. 1993; Rosgen et al. 1985), monitoring fish population response (Gorman and Karr 1978; Hayes et al. 1996; Zorn and Wiley 2006), and monitoring ongoing anthropogenic change such as diverse impacts of sedimentation (Hartman and Hakala 2006; Kaller 2001; Kaller and Hartman 2004; Lisle and Lewis 1992; Newcombe and MacDonald 1991; Waters 1995; Wood and Armitage 1997). Though examination of fish and habitat are sometimes separate undertakings, the management of stream fishes requires an understanding of the physical aspects of stream habitat (Kohler and Hubert 1999; Orth and White 1993; Wattage et al. 2005).

Understanding of benthic substrate has increased in the past three decades. Substrate knowledge has progressed from general statements concerning use and preferences, which often change during breeding cycles when stream fish seek out appropriately sized substrate, flow, and depth for spawning (Aadland 1993; Bjornn and Reiser 1991), to specific quantification of substrate deposition and size requirements by individual fish species and impacts caused by changes in substrate (Wu and Wang 2002). Anthropogenic and some natural occurrences of catastrophic substrate change such as hurricanes or floods may have devastating impact to fish communities within streams (Kaufman 1992; Matthews 1986; Scheffer et al. 2001; Scheffer and Carpenter 2003). Even minimal amounts of siltation, when occurring during spawning times, may alter fish abundance (Hartman and Hakala 2006; Lisle and Lewis 1992; Soulsby et al. 2001). Abnormal substrate alteration may be a symptom of other less localized problems in the watershed, such as farming practices, roads, or other development (Waters 1995; Wu and Wang

2002) including acidic conditions from acid precipitation and mining effluent often impair stream ecosystems in Appalachia (Likens et al. 2007; Wisniewski and Keitz 1983).

Monitoring and assessment of stream substrate is also an integral component for current management concerns. For example, invasive species often become more easily established in impaired ecosystems (Baxter et al. 2004; Byers 2002). Frequent stream assessment helps monitor ongoing habitat fragmentation which is important because habitat loss, such as filled interstitial spaces of benthic substrate, may increase vulnerability of benthic fishes to extirpation (Fagan et al. 2002). Species occurrence estimates may be created based solely on known habitat use and available habitat data such as substrate (Dunham et al. 2002; Lamouroux et al. 1999; Wildhaber and Lamberson 2004).

Water depth is an under-discussed stream habitat variable often used in conjunction with substrate and flow velocity in stream habitat assessments. Depth influences species occurrence seasonally; fish seek deep refuge during cold winter months and inhabit highly specific depth niches during spawning (Baltz et al. 1991). Depth also regulates the size of fishes in streams (Harvey and Stewart 1991). As riparian habitat is altered or removed, fish relationship to depth may change and institute fish community shifts (Jones III et al. 1999). Light penetration is a function of depth and water clarity and influences fish occurrence at any given location (Beeton 1958). Various aspects of remote sensing such as LiDAR, aerial, and satellite imagery also lose effectiveness as depth increases or as water clarity declines (Bustamante et al. 2009; Muirhead and Cracknell 1986). This is true particularly in absence of LiDAR band adjustment (Bustamante et al. 2009; Muirhead and Cracknell 1986). Depth measurements become more limited as turbidity increases when using LiDAR and remote sensing data (Collin et al. 2008).

As with depth, many fish species have specific flow velocity requirements at various stages in their life history (Hill et al. 1991). Methodologies of classification and measurement of flow for the benefit of fisheries, recreation, and industry have been developed due to importance of this variable (Rosenfeld 2003). Flow measurements are often included in stream studies, in part, to quantify minimum flow requirements for species (Aadland 1993). In spite of the hurdles presented by competing interests such as recreation, fisheries, and industry, legislation is frequently undergoing development to better preserve flow regimes and minimize damage to stream fish communities while still providing water source and flow for other uses (Acreman and Dunbar 2004; Boyd 2003; Reiser et al. 1989). Flow is also related spatially to substrate deposition and depth in streams, a condition which may be exploited to create and evaluate stream habitat models.

Spatial Modeling of Stream Variables

Sampling and spatial modeling of stream habitat data have developed alongside gains in computing power and use of geographic information systems (GIS; Fisher 2004). However, as analysis of stream spatial data has evolved, it has encountered the complex issues of spatial heterogeneity, non-stationarity, spatial autocorrelation, and multicollinearity. Spatial heterogeneity refers to the level of order and disorder of variables in their arrangement in geographical space of the stream. Spatial non-stationarity indicates the tendency of variables to move and change over the geographic space of the stream. Spatial autocorrelation is the condition of values of variables being irremovably connected to one another (e.g. a depth value always has an associated and possibly related flow value). Multicollinearity defines the quality of redundancy in variables, or variables which tell the same story in a model, inflating model accuracy. Such issues indicate need for more appropriate spatial statistical methodologies

(Austin 2007) as prior methods may no longer fully address needs of stream biologists (Poole et al. 1997). This has led to a decade long call to action for new methodology and focus within the fisheries community (Caddy 1999; Minns 2001; Poole et al. 1997; Stephenson and Lane 1995).

Some spatial data collection on streams has been transferred to remote sensing methodologies because of large, spatially explicit data requirements (Griffith 2002; Jensen 2009; Mertes 2002; Torgersen et al. 2001). By creating efficiencies in data collection, technologies such as Light Image Detection and Ranging (LiDAR), hydro-acoustics, high resolution aerial and satellite imagery of natural and hyper-spectral quality have allowed spatial analysis of streams and other aquatic systems with some success (Jensen 2009; Mertes 2002; Muirhead and Cracknell 1986; Valavanis et al. 2008).

Though impressive, remote sensing technologies have limitations and are not a panacea for aquatic assessments (Dubayah and Drake 2000). LiDAR is limited by the reflective surface of water and aerial imagery is limited by atmospheric scattering, weather, and season (leaf on/leaf off, ice; (Jensen 2009). During leaf off, aquatic vegetation, canyons and surface geology may impede adequate capture of stream information for spatial analysis (Jensen 2009; Mertes 2002). Steep gradients agitate water to an extent which aerial visual methods of collecting data are not effective, as may turbidity due to sediment load, or water chemistry.

As the trend moves towards understanding of aquatic habitat at various scales, tools such as GIS are developing as a framework towards this end (Smith et al. 2007; Store and Jokimäki 2003). Spatial qualities of data such as auto-correlation, non-stationarity, and scale are approached and addressed by these new tools to help analysis of fish habitat. Incorporation of spatial qualities of data have promoted deeper understanding of their populations and illustrated problems with prior analysis (González-Megías et al. 2005; Legendre 1993).

Spatial analysis of stream heterogeneity has also undergone intense growth (Cooper et al. 1997; Levins 1969; Palmer et al. 1997; Winemiller et al. 2010). Creation of maps and analysis of spatial data in streams are becoming staple products in government and non-government organization assessment regimes because of the importance to physical habitat assessment, management, and ability to provide more cost effective study structure (Bickers 2003; Gergel et al. 2007; Meaden 2001). While stream heterogeneity provides challenges for stream analysis and assessment, it is also an issue impeding accurate map creation and spatial variable analysis (Zhang and Gove 2005). Heterogeneity tends to cause excessive over and under estimation of habitat variables which limits map and prediction accuracies (Zhang and Gove 2005). Therefore, spatial heterogeneity in streams requires additional study to assess the impact it has on accuracy of mapping and assessment of stream habitat variables.

Previous studies have been made to quantify occurrence of spatial heterogeneity in streams (Cooper et al. 1997). However, the review by Cooper et al. (1997) deals with *functional heterogeneity* surrounding organism occurrence and not with *structural heterogeneity* directly. The studies and methodology examined tend to deal with the same topic (functional heterogeneity) and neither discuss heterogeneity's role from a standpoint of effective maps nor quantify levels of spatial heterogeneity. Cooper acknowledges that spatial heterogeneity studies have been limited due to a lack of tools (Cooper et al. 1997), however, newly available statistical tools in the form of spatially aware geographically weighted regression (GWR) have alleviated that issue (Fotheringham et al. 1996).

Geographically weighted regression is used in several chapters of this dissertation. It is a modified regression method for the analysis of spatial data with inherent qualities of spatial autocorrelation and non-stationarity which stream habitat variables in this study often display. More

specifically, GWR was created to model data with heterogeneity, which stream habitat variables often exhibit (Charlton et al. 2005; Fotheringham et al. 1996). The method was first introduced in the mid 1990's (Fotheringham et al. 1996), and later applied to ecological studies (Austin 2007; Kupfer and Farris 2007), and oceanic fisheries research and management (Wang et al. 2006; Windle et al. 2009). Researchers have examined and compared the applicability of GWR for analysis of spatial data relative to that of other regression methods (Ali et al. 2007; Gao et al. 2006; LeSage 2001), however, such analyses have not extended far into the field of fisheries science (e.g. Windle 2009).

Geographically weighted regression is often compared to OLS because it illustrates benefits of using a spatial non-stationarity approach in statistical models (Charlton et al. 2005; Fotheringham et al. 1996; Kupfer and Farris 2007; Lo 2008). The GWR method differs from that of OLS in several ways. The standard expression for OLS is $y_i = \beta_0 + \sum_k \beta_k x_{ik} + \varepsilon_i$ which is modified to include the expression of (u_i, v_i) , in both the β_0 and β_k facets of the GWR regression formula (Fotheringham et al. 1996). One may think of (u_i, v_i) , as an x,y coordinate. In effect, a coordinate location is added to each data observation or i and thus gives the new equation $y_i = \beta_0(u_i, v_i) + \sum_k \beta_k(u_i, v_i)x_{ik} + \varepsilon_i$ (Charlton et al. 2005; Fotheringham et al. 1996), which is useful in the analysis of spatial data (Mennis 2006). Further, GWR does not follow the assumption of homoskedacity, or static variance, but instead calculates a specific variance for data within a zone, or search radius of each predictor variable. Variance becomes dynamic and allows for a more accurate regression of data when non-stationarity (heteroskedacity) is present (Fotheringham et al. 1996; Zimmerman 2003). Geographically weighted regression calculates specific variance for each coordinate point. In this way, GWR may be considered a local rather

than a global model, because variance is not an averaged or single value. This is important for stream habitat models because variables such as substrate, depth and flow, often exhibit dynamic localized variation along habitat transition zones.

In addition to GWR, there are various interpolative methods including inverse distance weighted, ordinary and universal kriging, natural neighbor interpolation, point interpolation, trend, and spline interpolation to create predictions of stream habitat data (Childs 2004). Comparisons have shown that each method of interpolation has its own strengths in dealing with data of different types and number (Bennett et al. 1984; Kratzer et al. 2006; Le et al. 1997; MacKay 1992; Sambridge et al. 1995; Sheehan and Welsh 2009; Zimmerman et al. 1999).

In streams, natural neighbor interpolation has shown to be the most accurate data selection method for spatial analysis (Sheehan and Welsh 2009). Natural neighbor interpolation, used in this dissertation, works well with large datasets and has a nearly identical algorithm as inverse distance weighted interpolation. Further, natural neighbor interpolation is based on Thiessen polygon networks, and weights adjacent data within a specified search radius. Natural neighbor interpolation works well with stream habitat data such as substrate because depositional patterns in rivers are typically well ordered and not random (Jopling and Forbes 1979; Lunt et al. 2004; Purkait 2002). Depth, flow velocity, and substrate may have a high degree of spatial autocorrelation which further helps prediction accuracy (Kratzer et al. 2006; Smith and Ferguson 1995). Stream habitat variables of depth, flow velocity, and substrate have been predicted accurately using natural neighbor interpolation when applied to small amounts of data (Sheehan and Welsh 2009).

Measurement of the spatial distribution of data typically includes Moran's I, which may be used as a measure of heterogeneity (Anselin 2002; Anselin et al. 1996; Zhang and Gove

2005). Heterogeneity is also made up of two main components; complexity and variability (Li and Reynolds 1995). As discussed by Li and Reynolds (1995), a study may examine structural heterogeneity and functional heterogeneity. Structural heterogeneity focuses entirely on the structure of habitat, and not its effects on ecological function. Moran's I measures structural heterogeneity by looking at spatial distribution of data within a study area and returns values from -1, which indicate a fully dispersed spatial pattern, to 1 which indicates a clustered, ordered spatial pattern. A zero value indicates a random spatial pattern (Anselin et al. 1996; Moran 1953).

Summary Statement

To sum, this literature review examined the importance of assessing stream habitat data and why spatial attributes of those data are valuable. I also reviewed the types of analysis contained within the dissertation including GWR and natural neighbor interpolation. Further, I have identified the reasons behind their use, and why they may provide additional insight to stream habitat assessments and study of aquatic organisms. The chapters in this dissertation should be viewed like the river continuum concept- each aspect building upon aspects of the last (Vannote et al. 1980). Emphasis on stream habitat in this dissertation is centered around the reasoning that successful understanding of aquatic populations depends largely on availability, and thus accurate modeling and assessment, of appropriate habitat (Gergel et al. 2007; Orth and White 1999).

Literature Cited

Aadland, L. P. 1993. Stream habitat types: Their fish assemblages and relationship to flow. North American Journal of Fisheries Management 13(4):790-806.

- Acreman, M., and M. J. Dunbar. 2004. Defining environmental river flow requirements- a review. *Hydrology and Earth System Sciences* 8(5):861-876.
- Ali, K., M. D. Partridge, and M. R. Olfert. 2007. Can geographically weighted regressions improve regional analysis and policy making? *International Regional Science Review* 30(3):300.
- Anselin, L. 2002. Under the hood issues in the specification and interpretation of spatial regression models. *Agricultural Economics* 27(3):247-267.
- Anselin, L., A. K. Bera, R. Florax, and M. J. Yoon. 1996. Simple diagnostic tests for spatial dependence. *Regional Science and Urban Economics* 26(1):77-104.
- Austin, M. 2007. Species distribution models and ecological theory: a critical assessment and some possible new approaches. *Ecological Modelling* 200(1-2):1-19.
- Baltz, D. M., B. Vondracek, L. R. Brown, and P. B. Moyle. 1991. Seasonal changes in microhabitat selection by rainbow trout in a small stream. *Transactions of the American Fisheries Society* 120(2):166-176.
- Baxter, C. V., K. D. Fausch, M. Murakami, and P. L. Chapman. 2004. Fish invasion restructures stream and forest food webs by interrupting reciprocal prey subsidies. *Ecology* 85(10):2656-2663.
- Beeton, A. M. 1958. Relationship between Secchi disc readings and light penetration in Lake Huron. *Transactions of the American Fisheries Society* 87(1):73-79.
- Bennett, R. J., R. P. Haining, and D. A. Griffith. 1984. Review article: The problem of missing data on spatial surfaces. *Annals of the Association of American Geographers* 74(1):138-156.

- Bickers, A. N. 2003. Cost effective marine habitat mapping from small vessels using GIS, sidescan sonar and video. Coastal GIS. Australian Online Coastal Information, Geoscience Australia.
- Bjornn, T. C., and D. Reiser. 1991. Habitat requirements of salmonids in streams. American Fisheries Society Special Publication (19):83-138.
- Boyd, J. A. 2003. Hip deep: A survey of state instream flow law from the Rocky Mountains to the Pacific Ocean. Natural Resources Journal 43(Fall):1151-1216.
- Bustamante, J., F. Pacios, R. Díaz-Delgado, and D. Aragonés. 2009. Predictive models of turbidity and water depth in the Doñana marshes using Landsat TM and ETM+ images. Journal of Environmental Management 90(7):2219-2225.
- Byers, J. E. 2002. Impact of non-indigenous species on natives enhanced by anthropogenic alteration of selection regimes. Oikos 97(3):449-458.
- Caddy, J. F. 1999. Fisheries management in the twenty-first century: will new paradigms apply? Reviews in Fish Biology and Fisheries 9(1):1-43.
- Charlton, M., S. Fotheringham, and C. Brunsdon. 2005. Geographically weighted regression. NCRM Methods Review Papers (NCRM/006).
- Childs, C. 2004. Interpolating surfaces in ArcGIS spatial analyst. Pages 32-35 *in* ArcUser. ESRI, California.
- Collin, A., P. Archambault, and B. Long. 2008. Mapping the shallow water seabed habitat with the SHOALS. IEEE Transactions on Geoscience and Remote Sensing 46(10):2947-2955.
- Cooper, S. D., L. Barmuta, O. Sarnelle, K. Kratz, and S. Diehl. 1997. Quantifying spatial heterogeneity in streams. Journal of the North American Benthological Society 16(1):174-188.

- Cooper, S. D., S. Diehl, K. I. M. Kratz, and O. Sarnelle. 1998. Implications of scale for patterns and processes in stream ecology. *Austral Ecology* 23(1):27-40.
- Dubayah, R. O., and J. B. Drake. 2000. Lidar remote sensing for forestry. *Journal of Forestry* 98(6):44-46.
- Dunham, J. B., B. E. Rieman, and J. T. Peterson. 2002. Patch-based models to predict species occurrence: lessons from salmonid fishes in streams. Pages 327-334 *in* Predicting species occurrences: Issues of accuracy and scale. Island Press, Covelo, CA.
- Fagan, W. F., P. J. Unmack, C. Burgess, and W. Minckley. 2002. Rarity, fragmentation, and extinction risk in desert fishes. *Ecology* 83(12):3250-3256.
- Fisher, W. L. 2004. Future of geographic information systems in fisheries. Pages 259-266 *in* W. L. Fisher, and F. J. Rahel, editors. *Geographic Information Systems in Fisheries*. American Fisheries Society, Bethesda, MD.
- Fotheringham, A. S., M. Charlton, and C. Brunsdon. 1996. The geography of parameter space: an investigation of spatial non-stationarity. *International Journal of Geographical Information Science* 10(5):605-627.
- Frissell, C. A., W. J. Liss, R. E. Gresswell, R. K. Nawa, and L. Ebersole. 1997. *A resource in crisis: changing the measure of salmon management*. Chapman and Hall, New York, NY.
- Gao, X., Y. Asami, and C. J. F. Chung. 2006. An empirical evaluation of spatial regression models. *Computers and Geosciences* 32(8):1040-1051.
- Gergel, S., Y. Stange, N. Coops, K. Johansen, and K. Kirby. 2007. What is the value of a good map? An example using high spatial resolution imagery to aid riparian restoration. *Ecosystems* 10(5):688-702.

- Gido, K. B., J. A. Falke, R. M. Oakes, and K. J. Hase. 2006. Fish-habitat relations across spatial scales in prairie streams. Pages 265–285 *in* R. M. Hughes, L. Wang, and P. W. Seelbach, editors. American Fisheries Society Symposium 48. American Fisheries Society, Bethesda, Maryland.
- González-Megías, A., J. M. Gómez, and F. Sánchez-Piñero. 2005. Consequences of spatial autocorrelation for the analysis of metapopulation dynamics. *Ecology* 86(12):3264-3271.
- Gorman, O. T., and J. R. Karr. 1978. Habitat structure and stream fish communities. *Ecology* 59(3):507-515.
- Griffith, J. A. 2002. Geographic techniques and recent applications of remote sensing to landscape-water quality studies. *Water, Air, & Soil Pollution* 138(1):181-197.
- Hartman, K. J., and J. P. Hakala. 2006. Relationships between fine sediment and brook trout recruitment in forested headwater streams. *Journal of Freshwater Ecology* 21(2):215-230.
- Harvey, B., and A. Stewart. 1991. Fish size and habitat depth relationships in headwater streams. *Oecologia* 87(3):336-342.
- Hayes, D. B., C. P. Ferreri, and W. W. Taylor. 1996. Linking fish habitat to their population dynamics. *Canadian Journal of Fisheries and Aquatic Sciences* 53(1):383-390.
- Hill, M. T., W. S. Platts, and R. L. Beschta. 1991. Ecological and geomorphological concepts for instream and out-of-channel flow requirements. *Rivers* 2(3):198-210.
- Ian, M. 1999. The importance of physical habitat assessment for evaluating river health. *Fresh Water Biology* 41(2):373-391.
- Jensen, J. R. 2009. Remote sensing of the environment, volume 48. Prentice Hall, Upper Saddle River, NJ.

- Jones III, E., G. S. Helfman, J. O. Harper, and P. V. Bolstad. 1999. Effects of riparian forest removal on fish assemblages in southern Appalachian streams. *Conservation Biology* 13(6):1454-1465.
- Jopling, A. V., and D. L. Forbes. 1979. Flume study of silt transportation and deposition. *Geografiska Annaler. Series A, Physical Geography* 61(1/2):67-85.
- Jungwirth, M., S. Muhar, and S. Schmutz. 1995. The effects of recreated instream and ecotone structures on the fish fauna of an epipotamal river. *Hydrobiologia* 303(1):195-206.
- Kaller, M. D. 2001. Effects of sediment upon benthic macroinvertebrates in forested northern Appalachian streams. West Virginia University, Morgantown.
- Kaller, M. D., and K. J. Hartman. 2004. Evidence of a threshold level of fine sediment accumulation for altering benthic macroinvertebrate communities. *Hydrobiologia* 518(1):95-104.
- Karr, J. R., K. D. Fausch, P. L. Angermeier, P. R. Yant, and I. J. Schlosser. 1986. Assessing biological integrity in running waters. Pages 25 *in* A method and its rationale. Illinois Natural History Survey, Champaign, Special Publication. Illinois Natural History Survey, Champaign, IL.
- Kaufman, L. 1992. Catastrophic change in species-rich freshwater ecosystems. *BioScience* 42(11):846-858.
- Kohler, C. C., and W. A. Hubert. 1999. Inland fisheries management in North America, American Fisheries Society, Bethesda. MD, USA.
- Kratzer, J. F., D. B. Hayes, and B. E. Thompson. 2006. Methods for interpolating stream width, depth, and current velocity. *Ecological Modelling* 196(1-2):256-264.

- Kupfer, J., and C. Farris. 2007. Incorporating spatial non-stationarity of regression coefficients into predictive vegetation models. *Landscape Ecology* 22(6):837-852.
- Lamouroux, N., and coauthors. 1999. Predicting community characteristics from habitat conditions: fluvial fish and hydraulics. *Freshwater Biology* 42(2):275-299.
- Le, N. D., W. Sun, and J. V. Zidek. 1997. Bayesian multivariate spatial interpolation with data missing by design. *Journal of the Royal Statistical Society. Series B (Methodological)* 59(2):501-510.
- Legendre, P. 1993. Spatial autocorrelation: trouble or new paradigm? *Ecology* 74(6):1659-1673.
- LeSage, J. P. 2001. A family of geographically weighted regression models. Pages 241-263 *in* L. Anselin, R. J. Florax, and S. J. Rey, editors. *Advances in Spatial Econometrics*. Springer, Berlin, Germany.
- Levins, R. 1969. Some demographic and genetic consequences of environmental heterogeneity for biological control. *Foundation papers in landscape ecology* 15(3):237-240.
- Lewis, C., and coauthors. 1996. Considerations of scale in habitat conservation and restoration. *Canadian Journal of Fisheries and Aquatic Sciences* 53(S1):440-445.
- Li, H., and J. F. Reynolds. 1995. On Definition and Quantification of Heterogeneity. *Oikos* 73(2):280-284.
- Likens, G., S. C. Nodvin, W. Davis, and L. Zaikowski. 2007. Acid rain. C. J. Cleveland, editor *Encyclopedia of Earth*. Environmental Information Coalition, National Council for Science and the Environment.
- Lisle, T. E., and J. Lewis. 1992. Effects of sediment transport on survival of salmonid embryos in a natural stream: a simulation approach. *Canadian Journal of Fisheries and Aquatic Sciences* 49(11):2337-2344.

- Lo, C. P. 2008. Population estimation using geographically weighted regression. *GIScience & Remote Sensing* 45(2):131-148.
- Lunt, I. A., J. S. Bridge, and R. S. Tye. 2004. A Quantitative, Three-dimensional depositional model of gravelly braided rivers. *Sedimentology* 51(3):377-414.
- Mac Nally, R., and G. P. Quinn. 1998. Symposium introduction: The importance of scale in ecology. *Australian Journal of Ecology* 23(1):1-7.
- MacKay, D. J. C. 1992. Bayesian interpolation. *Neural Computation* 4(3):415-447.
- Matthews, W. J. 1986. Fish faunal structure in an Ozark stream: stability, persistence and a catastrophic flood. *Copeia* 1986(2):388-397.
- Meaden, G. J. 2001. GIS in fisheries science: foundations for the new millennium. Pages 3-29 *in* T. Nishida, P. J. Kailola, and C. E. Hollingsworth, editors. First International Symposium on GIS in Fishery Science. Fisheries Research Group, Saitama, Japan.
- Mennis, J. 2006. Mapping the results of geographically weighted regression. *The Cartographic Journal* 43(2):171-179.
- Mertes, L. A. K. 2002. Remote sensing of riverine landscapes. *Freshwater Biology* 47(4):799-816.
- Minns, C. K. 2001. Science for freshwater fish habitat management in Canada: current status and future prospects. *Aquatic Ecosystem Health & Management* 4(4):423-436.
- Montgomery, D. R., and J. M. Buffington. 1997. Channel-reach morphology in mountain drainage basins. *Bulletin of the Geological Society of America* 109(5):596.
- Montgomery, D. R., J. M. Buffington, and P. Timber-Fish-Wildlife. 1993. Channel classification, prediction of channel response, and assessment of channel condition. University of Washington, Seattle.

- Moran, P. A. P. 1953. The statistical analysis of the Canadian lynx cycle. I. Structure and prediction. *Australian Journal of Zoology* 1(1):163-173.
- Muhar, S., and M. Jungwirth. 1998. Habitat integrity of running waters—assessment criteria and their biological relevance. *Hydrobiologia* 386(1):195-202.
- Muirhead, K., and A. Cracknell. 1986. Airborne lidar bathymetry. *International Journal of Remote Sensing* 7(5):597-614.
- Newcombe, C. P., and D. D. MacDonald. 1991. Effects of suspended sediments on aquatic ecosystems. *North American Journal of Fisheries Management* 11(1):72-82.
- Orth, D. J., and R. J. White. 1993. *Stream Habitat Management*, Bethesda, Md.
- Orth, D. J., and R. J. White. 1999. Inland Fisheries Management in North America. Pages 249-283 *in* W. A. Hubert, and K. C.C., editors. *American Fisheries Society Bethesda, Maryland*, volume 2. American Fisheries Society, Bethesda, MD.
- Palmer, M. A., C. C. Hakenkamp, and K. Nelson-Baker. 1997. Ecological Heterogeneity in Streams: Why Variance Matters. *Journal of the North American Benthological Society* 16(1):189-202.
- Poole, G. C., C. A. Frissell, and S. C. Ralph. 1997. In-stream habitat unit classification: Inadequacies for monitoring and some consequences for management. *Journal of the American Water Resources Association* 33(4):879-896.
- Purkait, B. 2002. Patterns of grain-size distribution in some point bars of the Usri River, India. *India Journal of Sedimentary Research* 72(3):367-375.
- Reiser, D. W., T. A. Wesche, and C. Estes. 1989. Status of instream flow legislation and practices in North America. *Fisheries* 14(2):22-29.

- Rosenfeld, J. 2003. Assessing the habitat requirements of stream fishes: an overview and evaluation of different approaches. *Transactions of the American Fisheries Society* 132(5):953-968.
- Rosgen, D. L., and coauthors. 1985. A stream classification system. Pages 91-95 *in* First North American Riparian Conference, Rocky Mountain Forest and Range Experiment Station.
- Sambridge, M., J. Braun, and H. McQueen. 1995. Geophysical parametrization and interpolation of irregular data using natural neighbours. *Geophysical Journal International* 122(3):837-857.
- Scheffer, M., S. Carpenter, J. A. Foley, C. Folke, and B. Walker. 2001. Catastrophic shifts in ecosystems. *Nature* 413(6856):591-596.
- Scheffer, M., and S. R. Carpenter. 2003. Catastrophic regime shifts in ecosystems: linking theory to observation. *Trends in Ecology & Evolution* 18(12):648-656.
- Schlosser, I. J. 1990. Environmental variation, life history attributes, and community structure in stream fishes: implications for environmental management and assessment. *Environmental Management* 14(5):621-628.
- Scott, J., P. J. Heglund, and M. L. Morrison. 2002a. Predicting Species Occurrences. *Issues of Accuracy and Scale*. Island Press, Washington, DC.
- Scott, M. C., G. S. Helfman, M. E. McTammany, E. F. Benfield, and P. V. Bolstad. 2002b. Multiscale influences on physical and chemical stream conditions across Blue Ridge landscapes. *Journal of the American Water Resources Association* 38(5):1379-1392.
- Sheehan, K. R., and S. A. Welsh. 2009. An interpolation method for stream habitat assessments. *North American Journal of Fisheries Management* 29(1):1-9.

- Smith, A. D. M., E. J. Fulton, A. J. Hobday, D. C. Smith, and P. Shoulder. 2007. Scientific tools to support the practical implementation of ecosystem-based fisheries management. *ICES J. Mar. Sci.* 64(4):633-639.
- Smith, G. H. S., and R. I. Ferguson. 1995. The gravel-sand transition along river channels. *Journal of Sediment Research* 65(2a):423-430.
- Soulsby, C., A. F. Youngson, H. J. Moir, and I. A. Malcolm. 2001. Fine sediment influence on salmonid spawning habitat in a lowland agricultural stream: a preliminary assessment. *The Science of The Total Environment* 265(1-3):295-307.
- Stephenson, R. L., and D. E. Lane. 1995. Fisheries management science: A plea for conceptual change. *Canadian Journal of Fisheries and Aquatic Sciences* 52(9):2051-5056.
- Store, R., and J. Jokimäki. 2003. A GIS-based multi-scale approach to habitat suitability modeling. *Ecological Modelling* 169(1):1-15.
- Thompson, A. R., J. T. Petty, and G. D. Grossman. 2001. Multi-scale effects of resource patchiness on foraging behaviour and habitat use by longnose dace, *Rhinichthys cataractae*. *Freshwater Biology* 46(2):145-160.
- Torgersen, C. E., R. N. Faux, B. A. McIntosh, N. J. Poage, and D. J. Norton. 2001. Airborne thermal remote sensing for water temperature assessment in rivers and streams. *Remote Sensing of the Environment* 76(3):386-398.
- Townsend, C. R., S. Dolédec, R. Norris, K. Peacock, and C. Arbuckle. 2003. The influence of scale and geography on relationships between stream community composition and landscape variables: description and prediction. *Freshwater Biology* 48(5):768-785.
- Valavanis, V. D., and coauthors. 2008. Modelling of essential fish habitat based on remote sensing, spatial analysis and GIS. *Hydrobiologia* 612(1):5-20.

- Vannote, R. L., G. W. Minshall, K. W. Cummins, J. R. Sedell, and C. E. Cushing. 1980. The river continuum concept. *Canadian Journal of Fisheries and Aquatic Sciences* 37(1):130-137.
- Wang, L., P. W. Seelbach, J. Lyons, and R. H. Hughes. 2006. Effects of levels of human disturbance on the influence of catchment, riparian, and reach-scale factors on fish assemblages. Pages 199-219 *in* R. H. Hughes, L. Wang, and P. W. Seelbach, editors. *Landscape Influences on Stream Habitats and Biological Assemblages*. American Fisheries Society, Bethesda, MD.
- Waters, T. F. 1995. Sediment in streams. Sources, biological effects, and control. *American Fisheries Society Monograph* 21(2):203-217.
- Wattage, P., S. Mardle, and S. Pascoe. 2005. Evaluation of the importance of fisheries management objectives using choice-experiments. *Ecological Economics* 55(1):85-95.
- White, R. J. 1996. Growth and development of North American stream habitat management for fish. *Canadian Journal of Fisheries and Aquatic Sciences* 53(S1):342-363.
- Wildhaber, M. L., and P. J. Lamberson. 2004. Importance of the habitat choice behavior assumed when modeling the effects of food and temperature on fish populations. *Ecological Modelling* 175(4):395-409.
- Windle, M. J. S., G. A. Rose, R. Devillers, and M. J. Fortin. 2009. Exploring spatial non-stationarity of fisheries survey data using geographically weighted regression (GWR): an example from the Northwest Atlantic. *ICES Journal of Marine Science* 67(1):145-154.
- Winemiller, K. O. 2005. Life history strategies, population regulation, and implications for fisheries management. *Canadian Journal of Fisheries and Aquatic Sciences* 62(4):872-885.

- Winemiller, K. O., A. S. Flecker, and D. J. Hoeinghaus. 2010. Patch dynamics and environmental heterogeneity in lotic ecosystems. *Journal of the North American Benthological Society* 29(1):84-99.
- Wisniewski, J., and E. L. Keitz. 1983. Acid rain deposition patterns in the continental United States. *Water, Air, & Soil Pollution* 19(4):327-339.
- Wood, P. J., and P. D. Armitage. 1997. Biological effects of fine sediment in the lotic environment. *Environmental Management* 21(2):203-217.
- Wu, F.-C., and C.-F. Wang. 2002. Effect of flow-related substrate alteration on physical habitat: A case study of the endemic river loach, *Sinogastromyzon puliensis* (Cypriniformes, Homalopteridae) downstream of Chi-Chi Diversion Weir, Chou-Shui Creek, Taiwan. *River Research and Applications* 18(2):155-169.
- Zhang, L., and J. H. Gove. 2005. Spatial assessment of model errors from four regression techniques. *Forest Science* 51(4):334-346.
- Zimmerman, D. 2003. Geographically weighted regression. *Journal of the American Statistical Association* 98(463):765-767.
- Zimmerman, D., C. Pavlik, A. Ruggles, and M. P. Armstrong. 1999. An experimental comparison of ordinary and universal kriging and inverse distance weighting. *Mathematical Geology* 31(4):375-390.
- Zorn, T. G., and M. J. Wiley. 2006. Influence of landscape characteristics on local habitat and fish biomass in streams of Michigan's lower peninsula. Pages 375-393 *in* American Fisheries Society Symposium. American Fisheries Society, Bethesda, MD.

Chapter 2

An Effective Data Selection Pattern for Modeling Benthic Substrate in Streams; Using Ordinary Least Squares Residuals to Guide Data Collection

Abstract

Benthic substrate is frequently assessed and studied for the management of streams because of its ecological importance. The mapping of benthic substrate in streams is an important part of stream habitat assessments, but commonly used methods of data collection, such as random and grid sampling, do not always produce accurate maps. This study examines random and grid collection methods and compares them to non-random sampling methods for assessment of benthic substrate. We examined the ability of each sampling method to produce accurate maps of benthic substrate on four stream sites of varying size. Because ability to reproduce sampling methods is important, use of ordinary least squares regression residuals to guide and lend consistency to non-random data selection was also explored. Non-random data sampling resulted in the most accurate maps as indicated by exact coordinate match percentage of predicted data to actual site data and root mean square error values. This study demonstrated the ability of non-random data sampling to create more accurate maps of benthic substrate than grid and random sampling methods, which are often used in current assessment protocols for wadeable streams. We also show ordinary least squares regressions may be used as a guide to increase accuracy in non-random point sampling of benthic substrate. By illustrating the usefulness and consistency in non-random sampling methodology for benthic substrate in streams, we provide aquatic habitat managers insight into the accuracy of their current assessments, and an avenue for improved decision making capability.

Introduction

Benthic substrate data collected for stream ecological assessments are the foundation on which many lotic studies and inland fisheries management actions are dependent upon. Stream habitat assessment involves collection of benthic substrate information to help provide a comprehensive snapshot of stream ecological status. For this reason, the collection of benthic substrate data occurs across ecologic scales at varying levels of detail. For example, stream assessment protocols including the small streams wadeable assessment protocol, Regional Environmental Monitoring and Assessment Protocol (REMAP), and Rapid Visual Habitat Assessment (RVHA) call for random pebble counts to quantify substrate [1]. Benthic substrate assessment also aids in the identification of stream type [2-4]. However, recent developments in spatial analysis methods of aquatic ecologic data afforded by geographic information systems (GIS) have ushered in new requirements for the detail, type, and amount of data required for collection. The progression towards increasingly detailed spatial assessment of stream benthic substrate is influenced by the growing body of literature acknowledging the importance of spatial qualities in habitat data [5-9]. For stream habitat assessment and sampling of benthic substrate, this means the inclusion of inherent spatial qualities in addition to that which standard substrate sampling typically contains [10].

A large proportion of spatial data collection has been transferred to remote sensing methodologies because of large, spatially explicit data requirements. By creating efficiencies in data collection, technologies such as LiDAR, hydro-acoustics, high resolution aerial and satellite imagery of natural and hyperspectral quality have allowed spatial analysis of streams and other aquatic systems with some success. However, each of the aforementioned technologies has limitations and are not a panacea for aquatic assessments [11]. For instance, LiDAR is limited by

the reflective surface of water. Aerial imagery is limited by atmospheric scattering, weather, and season (leaf on/leaf off, ice).

Stream habitat data assessments present particular challenges to remote sensing, particularly those measuring benthic substrate. Even during leaf off, aquatic vegetation, canyons and surface geology may impede adequate capture of stream information for spatial analysis. Constant steep gradient may agitate water to an extent which aerial visual methods of collecting data are not effective. Streams in some regions maintain high levels of turbidity due to sediment load, or water chemistry which allows for algal blooms or other biologic process impacting water clarity. Because of these issues, on site, non-remote sensing data collection is still necessary for many aspects of stream habitat assessments including benthic substrate.

Identification of the best pattern to sample spatial data in the field is important because on site physical collection of habitat data is slow (in comparison to remote sensing), inefficient, and requires extensive work hours for its collection. In addition, physical data collection cannot gather the millions of points at high resolution which remote sensing accomplishes in a timely manner. Therefore, most field data collection is less complete and a form of parameterization (characterization based on qualities within set boundaries or classes of data) of the actual habitat.

Field data may be collected in three common spatial patterns for analysis; grid, random, and non-random. However, there is disagreement over which pattern of data collection is best suited to mapping and assessment of spatial data [12]. The answer may depend on the variable and what scale it is being regarded [13]. Point selection for random and grid sampling is relatively automatic once scale of collection is established. However, which data to select non-randomly along a streambed to best model substrate deposition is less easily decided. Clarification of this question with respect to benthic substrate collection would aid in accuracy of

assessments. Further, ability to reproduce selection methodology for assessment and mapping purposes is also an important consideration; while remote sensing methods are somewhat standardized and allow comparison of streams, in field, physical stream assessment involving spatial data are much less organized.

In the field, non-random point selection of benthic substrate may benefit from ordinary least squares regression (OLS). Residuals from OLS regression are of interest to point selection because they explicitly measure the amount of spatial error, positive and negative, associated with predictions of each data point within the model [14]. By observing OLS standard errors mapped in a GIS, it may be possible to establish with certainty, benthic substrate transition zones and thus more appropriately select and collect field data for stream habitat assessments. While strategies have been suggested, no methodology is described in detail defining the selection of key benthic substrate points.

Therefore, because actual field data collection of benthic substrate is a necessary part of stream habitat assessments, this study explores the impact to assessment accuracy and maps caused by data selected in random, grid, and non-randomly selected patterns. Two main questions are addressed in this study. First, is a non-random data selection process effective for modeling of benthic substrate in streams? Second, are residuals from non-spatial regressions useful in guiding spatial field data collection including benthic substrate? Further, use of residuals from OLS regressions may provide an easily repeatable spatial sampling pattern for benthic substrate assessments in streams. Specifically, the study addresses whether random, grid, or non-random sampling patterns are most effective for creation of maps and stream habitat assessments. These products are important to aquatic habitat managers, in part, because of an increased ability to make management decisions from more accurate stream habitat assessments.

Methods

Study Sites and Data Collection

Benthic substrate data were collected from four wadeable streams for this study. Two streams were located in West Virginia; one each in Monongalia and Kanawha Counties and two streams were located in the Greater Yellowstone Ecosystem, Gallatin National Forest. The Elk River site was located downstream of the effluent of Big Sandy Creek in Kanawha County, West Virginia (81°21'3.857"W, 38°29'20.73"N) and the site measured 27 meters long by 22 meters wide. The second eastern study site was located on Aarons Creek, which lies within the Monongahela River system in Monongalia County, West Virginia (79°56'0.625"W 39°37'8.69"N). The Aarons site measured 23.3 meters long by 8.7 meters wide. The first of our two western sites was located in the Gallatin National Forest on Little Wapiti Creek, Montana, (111°16'53"W, 45°2'20"N). The Wapiti Creek site measured 33.5 meters long by 10 meters wide. The second western study site was located in another tract of Gallatin National Forest on Grayling Creek, Montana (111°6'16"W, 44°48'16"N). The Grayling Creek site measured 27.5 meters long by 18 meters wide.

Study sites were delineated by grid cells (0.3 m resolution per cell) using a fifty meter tape measure, laser rangefinder, and flagging (later removed). Starting at the downstream left of each site, values for benthic substrate size and depth were recorded for each x,y coordinate. Substrate values were recorded along a continuous scale in millimeters from 0.05 to >300 mm based on intermediate axis diameter. This was repeated until the site was captured in a complete grid of x,y coordinate points. Thus, actual values of substrate size were recorded for each 0.1 m² cell for each study site. In ArcMap 10, corner points for each study site were geo-referenced and

exported to Microsoft Excel. Next, x,y coordinates were calculated for the remainder of cells in the site grid and appended to the initial dataset of water depth and substrate size. The final datasets were imported back to ArcMap 10 for analysis. Ordinary Least Squares (OLS) regression was run on each site using full resolution datasets (0.3 m). Resulting standard error residual maps of OLS regressions were converted into raster maps and reclassified into two categories; overestimation and underestimation of expected value (Figure 1).

Point Selection

In sum, four different point selection methods will be used to create comparable data layers. Point sampling will take place in grid, random, non-random guided by OLS residuals, and non-random guided by OLS residuals and actual substrate deposition patterns. Each selection pattern will be performed on all four study sites at cell resolution of one and two meters.

To create grid sampling points for comparison, one and two meter grids with points at the centroid of each grid cell were created over each study site using the create fishnet feature tool (data management tools). The amount of points in each grid was recorded and that amount was then used to create random points within the extent of the same site using the create random points tool (data management tools).

To select non-random points, site boundaries were defined in detail because natural neighbor interpolation (spatial analyst toolset), which would be used to create maps, predicts values only to the maximum extent of the inputs used for the model [15]. Previously reclassified OLS residual maps were set to 45% transparency and superimposed on the full resolution datasets for each site. Full datasets were not classified by substrate category to avoid point selection bias. Using the same number of points applied to the creation of grid and random point layers described previously, non-random points were selected based on the location of the

boundary between over and underestimation on the OLS residual maps. To maximize potential accuracy for each non-random map, features were outlined by selecting points from small to large scale. The easternmost, westernmost, southernmost, and northernmost boundaries of a patch of OLS overestimation were selected first. If points were still available after major features were outlined in this manner a secondary round of points was applied around the perimeter of each feature to gain additional accuracy. This was done until the amount of allotted points for both grids were complete at each of the four study sites.

The final set of non-random point selection layers was created using a combination of OLS residuals and actual substrate deposition. Points for these layers were selected using the same procedure as outlined for OLS regression residual point selection. In addition, some points were selected based on the actual coverage, which would not have been selected without looking at the substrate deposition pattern. As an example of layers created for each of four sites, the Little Wapiti site had a total of 201 points for its one meter grid, and a random points layer, OLS non-random layer, and OLS and substrate non-random layer were then created within the extent of the Little Wapiti site also with 201 points (Figure 2, Table 1).

Analysis

Natural neighbor interpolation was performed on all random, grid, and non-random data layers. Natural neighbor is an appropriate interpolation method for substrate because it weights values closest to it more heavily when calculating a new value, similar to benthic substrate deposition pattern. Natural neighbor interpolation has shown promise in its ability to model stream variables in comparison to other types of interpolation for benthic substrate [12, 16]. Maps for visual comparison of substrate were created for all sites using the same substrate size scale (0 to >300mm). Predictive values from resulting interpolations were then appended to the

initial full dataset by extracting values to points for all layers. All data were exported for statistical analysis.

Accuracy of interpolations was demonstrated by calculating the percent of predicted values from OLS regressions which fell into the same substrate size class at each x,y coordinate location. Our substrate classification was as follows; boulders > 250mm, cobble 76-250mm, gravel 2-76mm, sand 0.25-2mm, and silt < 0.25mm. Interpolation accuracy was also calculated for each substrate size class and for all size classes combined for a complete image of accuracy. Root mean square error was also calculated for each site and compared RMSE from predictions. To calculate RMSE we established standard deviation for our sample and then used the formula $\frac{\sigma}{\sqrt{n}}$ where σ = standard deviation of the sample and n = sample size. Root mean square error was used because it indicates dispersion of data. Comparing dispersion levels of interpolations is another indicator of which interpolation matches best with the digital representation of our study site's substrate values.

Results

Non-random point selection guided by OLS regression residuals and substrate deposition pattern outperformed grid and random sampling patterns (Table 1). Non-random data selection maintained the most accurate maps based on exact coordinate match percentages when compared to the actual benthic substrate coverage of sites. This means interpolations using non-random data selection most accurately placed substrate of the correct type into the correct location, creating the most accurate maps of benthic substrate. This result was maintained for both one and two meter resolution samples (Table 1).

Grid, random, and non-random sampling patterns returned a range of 38.6% to 83.5% for one meter based sample interpolations and 32.6% to 77.9% for two meter based sample interpolations. Interpolations created with point selection guided only by OLS residual patterns were nearly as accurate as those using OLS residuals and the benthic substrate deposition pattern itself. The least accurate and most variable method of point selection pattern was random; random point selection created the greatest range in accuracy across sites from 32.94% to 79.10%. Grid patterns performed consistently in the middle range of all models. The percent difference between OLS residual sampling and OLS residuals and substrate deposition sampling pattern was slight, averaging 2.83%. Observation of accuracy for individual substrate type explains this result.

The following observations were made from visual inspection of maps: Loss in the range of data (maximum and minimum values recorded) and area of site lost to location of points was most prevalent in random and grid maps. Random and grid sampling pattern maximum and minimum values removed an average of 15mm from the range. The majority of loss occurred at the upper end of the boulder scale range. Non-random sampling method maps best maintained sight shape integrity and area (Figures 3-6). Superimposed under each data layer is the outline of the actual wetted width of each study site (Figures 3-6). Looking at the actual site boundary, and that applied by random, grid, and non-random sampling shows that random and grid sampling limits habitat boundaries. Non-random data sampling methodology more adequately captured the boundaries of the actual site.

Non-random, actively selected sampling patterns returned RMSE values closest to actual coverage for all sites (Table 2). Differences were typically twice as large between RMSE values of random and grid sampling patterns compared to actual coverage and those of non-random

OLS residual map and substrate guided sampling patterns. As with site interpolations, the Aarons Creek site had the least accurate RMSE values, while the Elk River site returned the closest to RMSE values to actual coverage.

Discussion

Based on this study, non-random data sampling patterns create more accurate maps of benthic substrate than either random or grid sampling methodology. While effective, grid and random data selection did not achieve the level of prediction accuracy shown by non-random benthic substrate data selection. Identification of the most effective spatial sampling pattern is important to aquatic habitat managers because it provides more accurate stream habitat assessments from which to base management decisions. The second conclusion drawn from this study is that OLS regression residuals created from field data may be used to effectively guide future benthic substrate data sampling for assessments and map making on the same stream. This is accomplished by mapping the residuals of OLS regression and using reclassifications of those maps to identify the line of transition between positive and negative deviation from actual values. As seen in study site maps, zones of actual substrate transition occur frequently within streams and OLS regression residuals explicitly define the important benthic substrate transition zones, thereby removing the guesswork associated with their identification.

Combining the repeatable nature of OLS regressions with the non-random point selection process for stream benthic substrate creates a standardized method. Stream habitat assessment methodology must be consistently repeatable to have value to aquatic habitat managers; preparing a map of OLS residuals from spatial data is a standard procedure in ArcMap 10 and is easily repeatable. Further, standardization of data selection in assessments allows for comparison

between sites with different substrate deposition pattern, as seen by the results and maps of benthic substrate in this study. By selecting points immediately besides the OLS transition border, the substrate values of greater spatial importance are selected which may have otherwise been ignored. As shown by mapped results, ordinary least squares residuals tend to trace zones of benthic substrate transition (Figures 3-6). There is some deviation from this pattern and the border between under and over estimation strays from actual benthic substrate transition zones. The areas which deviate from actual substrate deposition patterns are less intuitive, and their identification is important to the selection of non-random points because they would not have been selected by random, grid or substrate transition zone identification methods. Their addition allows pattern selection using a combination of substrate deposition pattern and OLS residuals to achieve greater accuracy.

The importance of small gains in model accuracy for mapping aquatic stream habitat data is also found directly within this study. By breaking down the model into substrate types, we observe large gains in accuracy for less frequently occurring substrates. Accuracy gains for benthic substrate found in less volume at a site are not easily discerned by measuring overall sight accuracy increases alone. Silt, which occurred infrequently and never occurred in more than 1% of the total substrate found at any site, could be predicted with zero and 100 percent accuracy, yet not affect the overall model accuracy by more than $\pm 0.5\%$. A single habitat variable which spatially dominates the total area of the site will result in models with higher average accuracy. The majority of the Elk River site was covered by sand, and predicting sand effectively elevated model accuracy even in models which failed to account for other substrates. In many cases, it is periphery habitat variables which play an important role in supporting the full diversity of a stream community [17, 18]. Models based on random or grid sampling, which

our results indicate a tendency for failure to identify both full range of habitat available and physical dimensions of the study reach in a stream, supply additional potential for incomplete assessments. Ability of non-random data selection to show presence of the full range of substrate and its values within sites adds value to the method. Visual inspection of maps also indicates that site boundaries were less effectively mapped in grid and random sampling methods. Analysis based on less than the full size of the wetted width of the stream provides information on a smaller area; the smaller the area sampled, the less likely a method is to catch unique habitats. Because edge boundaries are most often the area left out of spatial modeling using grid and random sampling, important edge habitats such as riparian area would be inadequately measured.

Some stream habitat assessments attempt to provide a comprehensive view of the available aquatic habitat. Following this directive, the best sampling method for stream habitat assessments provides the most complete picture of the variables being studied. Though no sampling methodology is infallible, those with tendency to supply mapping accuracy for a portion of the region assessed while inaccurately measuring others are not desirable. As demonstrated by the results of this study, grid and non-random spatial sampling patterns trend towards this type of behavior. This is an important point with concern to grid sampling, which is highly consistent in its data sampling methodology, but less consistent with its results within our study. Non-random OLS residual and benthic substrate deposition sampling patterns were more consistent in identifying the full range of values within each site.

In addition to inability of grid and random sampling to capture full range of benthic substrate values, data loss is also expressed by failure of grid and random sampling to capture the actual shape of study sites. Natural neighbor provides a conservative view of a site because it does not extend its reach beyond the spatial boundaries of the data. Other types of interpolation

may have their predictive extents extended beyond data collection points to a specified boundary, however, predictions in those extra-extent areas would be suspect regardless of the method used. While there is no control over this outcome with random and grid sampling besides increasing the amount of data collected, non-random data collection at any scale or size of dataset can limit this problem because of active selection of the full data range. Ground verification of presence should be required to confirm the full range of habitat variables in the model to guide its accuracy. Failure to accurately identify the presence of all types of habitat in non-random sampling would limit the ability of aquatic habitat managers to assess a stream because they would be missing valuable information.

The existence and identification of spatially important zones where a model may have the most intense data requirements are also important aspects of this study. By creating more accurate substrate models using OLS residuals and substrate deposition patterns we have demonstrated that some data are more spatially important. Grid and random data sampling assign equal importance to any spatial location. We see that our initial OLS regression maps create a unique spatial pattern separate than substrate deposition patterns. The non-random data selection method therefore more efficiently uses spatial information. Further, applying spatial information to stream habitat analysis may correct mistakes perpetuated without accounting for spatial information in a dataset [19]. This is a valuable outcome because ability to recognize spatially important habitat may help guide future assessments and management.

The connection between habitat availability (which this sampling pattern makes more accurate) and fish species occurrence and distribution is well documented. Therefore, ability to more effectively quantify area and location of available habitat through non-random data selection using OLS residuals as a guide allows aquatic scientists to make better inferences

concerning aquatic species population occurrence, health, and distribution within a stream. For instance, if one were to apply this method of benthic substrate modeling to estimate and map the availability of spawning substrate, stream habitat managers could actively assess and monitor change to its amount and location within the stream over time in relation to young of year reproductive success.

To sum, this study demonstrates the ability of non-random data sampling to create more accurate maps than grid and random sampling methods, which are often used in current assessment protocols for wadeable streams. We also show ordinary least squares regressions may be used as a guide to increase accuracy in non-random point sampling of benthic substrate. This has implications to other aquatic variables which are modeled spatially because the same method may potentially be used to guide their collection and mapping as well. By illustrating the usefulness and consistency in non-random sampling methodology for benthic substrate in streams, we provide aquatic habitat managers insight into the accuracy of their current assessments, and an avenue for improved data collection to aid in decision making capability.

Acknowledgments

We thank Chris Horn for input during the editing processes. Reference to trade names does not imply endorsement of commercial products by the U.S. government.

Literature Cited

1. Barbour, M.T., et al., *Rapid bioassessment protocols for use in streams and wadeable rivers*, U. EPA, Editor 1999, EPA: Washington.
2. Rosgen, D.L., *A classification of natural rivers*. Catena, 1994. **22**(3): p. 169-199.
3. Montgomery, D.R. and J.M. Buffington, *Channel-reach morphology in mountain drainage basins*. Bulletin of the Geological Society of America, 1997. **109**(5): p. 596.
4. Montgomery, D.R., J.M. Buffington, and P. Timber-Fish-Wildlife, *Channel classification, prediction of channel response, and assessment of channel condition*, 1993, University of Washington: Seattle. p. 110.
5. Torgersen, C.E., et al., *Spatial identification of tributary impacts in river networks*, in *River confluences, tributaries and the fluvial network*, C. Wood, K. Littleton, and K. Sheehy, Editors. 2008, Jon Wiley and Sons: West Sussex, England. p. 159-181.
6. Mykrä, H., J. Heino, and T. Muotka, *Scale related patterns in the spatial and environmental components of stream macroinvertebrate assemblage variation*. Global Ecology and Biogeography, 2007. **16**(2): p. 149-159.
7. Webb, B.W., et al., *Recent advances in stream and river temperature research*. Hydrological Processes, 2008. **22**(7): p. 902-918.
8. Clark, J.S., et al., *Spatial distribution and geomorphic condition of fish habitat in streams: an analysis using hydraulic modelling and geostatistics*. River Research and Applications, 2008. **24**(7): p. 885-899.
9. Brind'Amour, A., et al., *Multiscale spatial distribution of a littoral fish community in relation to environmental variables*. Limnology and Oceanography, 2005. **5**(2): p. 465-479.

10. Kraft, C.E. and D.R. Warren, *Development of spatial pattern in large woody debris and debris dams in streams*. Geomorphology, 2003. **51**(1-3): p. 127-139.
11. Dubayah, R.O. and J.B. Drake, *Lidar remote sensing for forestry*. Journal of Forestry, 2000. **98**(6): p. 44-46.
12. Atkinson, P.M., *Optimal sampling strategies for raster-based geographical information systems*. Global Ecology and Biogeography Letters, 1996. **5**(4/5): p. 271-280.
13. Fausch, K.D., et al., *Landscapes to riverscapes: bridging the gap between research and conservation of stream fishes*. BioScience, 2002. **52**(6): p. 483-498.
14. Charlton, M., S. Fotheringham, and C. Brunson, *Geographically weighted regression*. NCRM Methods Review Papers, 2005(NCRM/006).
15. Fortin, M.J., M.R.T. Dale, and J. ver Hoef, *Spatial analysis in ecology*. Encyclopedia of Environmetrics, 2006. **4**: p. 2051–2058.
16. Sheehan, K.R. and S.A. Welsh, *An interpolation method for stream habitat assessments*. North American Journal of Fisheries Management, 2009. **29**(1): p. 1-9.
17. Naiman, R.J., et al., *The potential importance of boundaries of fluvial ecosystems*. Journal of the North American Benthological Society, 1988. **7**(4): p. 289-306.
18. Borchardt, D., *Effects of flow and refugia on drift loss of benthic macroinvertebrates: implications for habitat restoration in lowland streams*. Freshwater Biology, 1993. **29**(2): p. 221-227.
19. González-Megías, A., J.M. Gómez, and F. Sánchez-Piñero, *Consequences of spatial autocorrelation for the analysis of metapopulation dynamics*. Ecology, 2005. **86**(12): p. 3264-3271.

Table 1. Root mean square values for each sampling method for all sites. Values in bold are closest to the actual coverage for each site.

Point Sampling Method	Elk	Aarons	Little Wapiti	Grayling
<i>Actual Coverage</i>	<i>0.448</i>	<i>2.814</i>	<i>2.082</i>	<i>1.301</i>
1 Meter Grid	0.277	1.892	1.355	0.999
2 Meter Grid	0.324	2.112	1.220	0.935
Random 1	0.314	1.876	1.350	0.933
Random 2	0.235	1.988	1.040	0.963
Residual Guided 1	0.409	2.067	1.561	1.105
Residual Guided 2	0.488	1.909	1.502	0.805
Res. & Sub Guided 1	0.403	2.390	1.911	1.484
Res. & Sub Guided 2	0.313	1.956	1.527	1.334

Table 2. Sampling information and percent of each substrate type predicted successfully at each site by sampling method. Percentages represent exact spatial (x,y coordinate) match. Values in bold indicate the best performing model. Non-random sampling guided by OLS residuals and substrate deposition pattern was the best performing sample pattern.

Substrate/Sample Info.	Elk	Aarons	Little Wapiti	Grayling	Elk	Aarons	Little Wapiti	Grayling
# of Sample Points Used	581	153	201	485	145	36	49	122
Sample Points Available	6192	1869	1901	4288	6192	1869	1901	4288
% of Site Used to Create Maps	9%	8%	11%	11%	2%	2%	3%	3%
	Residuals and Substrate 1				Residuals and Substrate 2			
Silt	0.00%	91.00%	16.85%	16.67%	0.00%	0.00%	0.00%	0.00%
Sand	77.19%	53.05%	8.51%	34.33%	68.22%	70.97%	0.00%	5.24%
Gravel	95.60%	28.47%	53.50%	89.74%	97.30%	22.70%	46.50%	93.76%
Cobble	63.16%	78.96%	86.89%	88.45%	33.33%	83.94%	83.15%	75.07%
Boulder	55.56%	44.44%	45.86%	34.78%	15.56%	14.03%	8.92%	15.42%
% of Site Predicted Correctly (proportional)	83.54%	57.97%	66.62%	80.24%	77.89%	48.44%	57.64%	73.72%
	OLS Residuals 1				OLS Residuals 2			
Silt	0.00%	85.71%	11.96%	0.00%	50.00%	8.57%	6.98%	0.00%
Sand	74.81%	32.43%	19.15%	29.03%	65.66%	27.31%	0.00%	3.23%
Gravel	95.91%	41.67%	61.40%	87.56%	85.89%	24.31%	41.65%	89.09%
Cobble	48.25%	93.91%	87.13%	90.13%	32.46%	88.52%	80.55%	85.50%
Boulder	57.78%	24.40%	19.11%	32.74%	31.11%	10.71%	8.33%	2.56%
% of Site Predicted Correctly (proportional)	81.97%	52.70%	66.21%	79.04%	72.25%	42.45%	55.35%	73.48%
	Random 1				Random 2			
Silt	0.00%	33.33%	3.90%	0.00%	0.00%	0.00%	3.64%	0.00%
Sand	67.49%	17.06%	6.38%	29.80%	66.04%	0.30%	0.00%	16.33%
Gravel	96.43%	8.25%	56.78%	91.08%	98.25%	30.95%	71.54%	65.56%
Cobble	27.66%	92.70%	88.93%	92.85%	4.55%	80.19%	59.83%	89.47%
Boulder	2.63%	19.91%	3.25%	4.69%	0.00%	4.09%	0.00%	6.35%
% of Site Predicted Correctly (proportional)	76.40%	45.78%	63.57%	79.10%	75.97%	32.94%	52.49%	65.70%
	Meter Grid 1				Meter Grid 2			
Silt	0.00%	0.00%	7.14%	0.00%	0.00%	0.00%	0.00%	0.00%
Sand	68.36%	6.51%	6.38%	23.68%	64.77%	6.25%	0.00%	17.24%
Gravel	88.21%	25.93%	57.85%	76.08%	93.05%	25.93%	61.33%	75.46%
Cobble	70.59%	90.35%	89.58%	93.70%	28.24%	90.75%	80.13%	86.84%
Boulder	2.56%	11.67%	3.85%	6.79%	0.00%	11.67%	1.99%	3.50%
% of Site Predicted Correctly (proportional)	75.10%	38.62%	64.42%	72.25%	73.16%	38.69%	61.43%	68.88%

Figure 1. Substrate deposition patterns (a) and ordinary least squares regression patterns of over and under estimation of substrate size (b) for Elk, Little Wapiti, Grayling, and Aarons sites. The boundary of over and under estimation closely follows the edge boundaries of substrate transitions zones.

Figure 2. Example on the Little Wapiti site of the the four point selection types used for interpolations.

Figure 3. Natural neighbor interpolations of Little Wapiti created by grid, random, and actively selected points. Maps illustrate the ability of ordinary least squares regression to identify substrate transition zones and aid in the point selection process allowing the method to outperform grid and random methods using the same amount of data.

Figure 4. Natural neighbor interpolations of Grayling Creek data created by grid, random, and actively selected points.

Figure 5. Natural interpolations of Aarons Creek data created by grid, random, and non random points. Aarons Creek maintained the most equal distribution of substrate of all sites.

Figure 6. Natural neighbor interpolations of Elk River data created by grid, random, and non-random data sampling. Elk River was the least heterogeneous site and was made up of more than 50% sand.

Substrate Deposition and Standard Error of Residuals

Point Sampling Methods Data Layers Created for Interpolation

Little Wapiti

Grayling

Aarons

Elk

Grid

Random

Non-random
OLS Residual
Sampling Non-random
OLS Residual
& Substrate
Assisted
Sampling

Actual Coverage

Substrate Size Categories
mm at Intermediate Axis

Chapter 3

Accuracy of stream habitat interpolations across spatial scales

Abstract

Stream habitat data are often collected across spatial scales because relationships among habitat, species occurrence, and management plans are linked at multiple spatial scales. Unfortunately, scale is often a limiting factor to gaining insight from spatial analysis of stream habitat data. Considerable cost is often expended to collect data at several spatial scales to gain accurate evaluation of spatial relationships in streams. To address utility of single set of data when used at varying scales, accuracy loss, and potential to lower data collection needs in stream habitat studies, we examined the influence of scale on accuracy of natural neighbor predicted depth, flow, and substrate maps. We measured two streams at 0.1 m² cell size over an area of 797 m² to create baseline for natural neighbor interpolated maps at 12 incremental scales ranging from a raster size of 0.1 m² to 13.38 m². Predictive maps exposed a logarithmic linear decay in r^2 and RMSE values for map accuracy for variables as scale departed from the original. Proportional accuracy of models was maintained up to 78% at scales 11 times more coarse than the original collection scale. Therefore, accuracy retention was suitable for assessment and management purposes at scales many times removed from the data collection scale. Our study is relevant to spatial modeling, fish habitat assessment, and stream habitat management because it highlights the potential of a single dataset to fulfill analysis needs otherwise requiring several scaled datasets at an increased cost of time and money.

Introduction

Stream habitat data at varying spatial scales provides integral information for their management and broad ecologic study. Typically, stream data are collected at multiple spatial scales to provide more complete representation of habitat and allow additional ecologic insight and analysis power. The spatial scale of stream habitat data is largely important due to connectivity among habitat patch dynamics, species occurrence, and life history [1-3]. Because of ecological links between scales, spatial analysis in varying forms has become a staple tool for examining multi-scale stream habitat data [1, 4, 5]. Collection of stream variables at multiple scales is also necessary for complex analysis of macroinvertebrates, fish habitat relationships, ecological processes, and stream habitat [1, 5-7]. While necessary for complex analysis, collection of stream habitat data at multiple scales is expensive and time consuming. Further, data at various scales has long presented problems such as pattern analysis and combining data at varying scales [8-10]. Interpolation methods represent a family of spatial statistics which are able to create products at multiple scales to aid in analysis and presentation of spatial data [4, 11, 12]. Natural neighbor interpolation has shown promise in producing practical maps of streams from small amounts of spatial data as shown by Sheehan and Welsh [12]. Demonstration of natural neighbor interpolation to accurately move between scales from a single stream habitat dataset may make multiple scale data collection redundant and provide an opportunity for cost and time savings.

Several data sets are often required for spatial analysis due to an inability of either dataset to be scaled for comparative purposes. Inability to make inferences, at scales other than that at which data were collected, relates directly to an unknown amount of accuracy loss when scaling between micro and macro scale stream habitat. As stated by Fisher and Rahel [13], data analysis

may only be as accurate as the largest scale of data collected. This leads stream biologists to collect data at the largest scale possible for each study. Unfortunately, there is an inverse relationship between the spatial scale of data and cost to acquire it; the finer the data scale required, the smaller the area able to be examined for a given amount of funding. Spatial interpolation, a family of statistical methods used in many GIS systems, can create continuous surfaces from spatial data for analysis purposes. Interpolation provides predictive values of variables in regions which have no data by using information from adjacent regions. This ability provides potential to use datasets at different scales than that which they were initially collected.

There are various interpolative methods including inverse distance weighted, several forms of kriging, natural neighbor, point interpolation, trend, and spline to create predictions of stream habitat data [14]. Many of these methods have been directly compared. Comparisons have shown that each method of interpolation has its own strengths in dealing with data of different types and number [12, 15-20].

Specifically, when selecting stream habitat data variables of depth, flow velocity, and benthic substrate at known locations, natural neighbor interpolation has shown to be the most accurate [12]. Natural neighbor works well with large datasets, and has a nearly identical algorithm as inverse distance weighted interpolation. Natural neighbor interpolation is based on Thiessen polygon networks, and weights adjacent data within a specified search radius. It takes a set of spatially located points, and creates a grid (raster map) of the area based on the input points at the centroid of each cell. Natural neighbor interpolation works well with stream habitat data such as substrate because depositional patterns in rivers are typically well ordered, and not random [21-23]. Depth, flow velocity, and substrate have a high degree of spatial autocorrelation which further helps prediction accuracy [16, 24]. While stream habitat variables of

depth, flow velocity, and substrate have been recreated accurately using natural neighbor interpolation when applied to small amounts of data [12], there has been no evaluation of the role of spatial scale on stream habitat model accuracy with this type of interpolation.

This study evaluates accuracy loss of predictive stream models when moving between scales using natural neighbor interpolation. We hypothesize that accuracy retention will be high enough to create practical data and maps for analysis purposes at scales well removed from the initially collected data scale. In relation to the hypothesis, the objective of the study was to examine accuracy of predictive models at stream sites using data on water depth, water velocity, and benthic substrate at multiple spatial scales. This study will help establish potential for the use of a single dataset across scales in stream habitat modeling. To our knowledge, there has been no such study on scalability of stream habitat data when using natural neighbor interpolation. Our study is relevant to spatial modeling, fish habitat assessment, and stream habitat management because it examines the potential of a single dataset to fulfill analysis needs which would otherwise require multiple datasets at varying spatial scales at an increased cost of time and money. Further, this study emphasizes the rate of accuracy loss between data scales while creating visual maps of stream habitat, which could potentially aid and streamline both data and stream habitat management.

Methods

Study Sites and Data Collection

Benthic substrate data were collected from four wadeable streams for this study. Two streams were located in West Virginia; one each in Monongalia and Kanawha Counties and two streams were located in the Greater Yellowstone Ecosystem, Gallatin National Forest. The Elk

River site was located downstream of the effluent of Big Sandy Creek in Kanawha County, West Virginia (81°21'3.857"W, 38°29'20.73"N) and the site measured 22 meters wide, by 27 meters long. The second eastern study site was located on Aarons Creek, which lies within the Monongahela River system in Monongalia County, West Virginia (79°56'0.625"W 39°37'8.69"N). The Aarons site measured 23.3 meters long by 8.7 meters wide. The first of our two western sites was located in the Gallatin National Forest on Little Wapiti Creek, Montana, (111°16'53.546"W, 45°2'20.639"N). The Wapiti Creek site measured 33.5 meters long by 10 meters wide. The second western study site was located in another tract of Gallatin National Forest on Grayling Creek, Montana (111°6'16.407"W, 44°48'16.878"N). The Grayling Creek site measured 27.5 meters long by 18 meters wide.

Study sites were delineated by grid cells (0.3 m² resolution per cell, an area of 0.1 m²) using a fifty meter tape measure, laser rangefinder, and flagging (later removed). A single piece of rebar was inserted into the bank material on each stream bank and high tensile line was secured to the rebar to guide the tape measure. As each row of data collection was finished the rebar was repositioned upstream to provide support for the next. Starting at the downstream left of each site, values for benthic substrate size and depth were recorded for each x,y coordinate. Substrate was recorded along a continuous scale in millimeters from 0.05 to >300 mm based on longest axis diameter. This was repeated until the site was captured in a complete grid of x,y coordinate points. Thus, actual values of substrate size were recorded for each 0.1 m² cell for each study site. In ArcMap 10, corner points for each study site were geo-referenced and exported to Microsoft Excel. Next, x,y coordinates were calculated for the remainder of cells in the site grid and appended to the initial dataset of water depth and substrate size. The final datasets were imported back to ArcMap 10 for analysis (Figures 1-4).

Data sets were created based on values at the centroid of each raster for 12 scales (including the base scale). Scale increments ranged from large, 0.1 m^2 , to smaller, 13.38 m^2 per cell (Table 1). Natural neighbor interpolation was run on each scale including the base scale to create maps for analysis. For each dataset, Ordinary Least Squares (OLS) regressions (ArcToolbox, ArcMap 10) were run to provide r^2 values and residuals for each x,y coordinate. Moran's I was used to test for significant levels ($p < 0.05$) of spatial autocorrelation in available data and residuals created from regressions [25]. Aforementioned maps of depth, flow velocity, and substrate residuals were created to display positive and negative prediction trends in the form of standard deviation at each x,y location. Null predictive values (found at site edge boundaries) were removed to allow clarified map view.

In this analysis, natural neighbor maps served as a visual and statistical base for scale accuracy comparisons because they have been shown to predict stream habitat data variables depth, and substrate well [12]. Each separate scale was compared to the base scale. Trend curves indicate natural neighbor interpolations (maps) of data collected at 0.1 m^2 are 100% accurate when a grid of all available data is used from the same resolution thus allowing the base scale interpolation to be used for effective comparison [12]. All predicted habitat values were exported at each scale of map and then were compared to observed values at the largest scale (Table 1). Root mean square error was calculated for comparison of predictive maps and baseline maps. Maps were created for each scale interpolation to aid analysis and provide important visual comparison of levels of accuracy loss between scales.

Root mean square error (RMSE) values were calculated for each scale, plotted, and appropriate trend lines applied. Plotting of RMSE values for each scale shows decay of accuracy for each scale effectively and compliments r^2 values comparing accuracy of scales 2-12 to the

base scale because RMSE decreases as proportional r^2 increases. Values of r^2 on RMSE graphs indicate trend line fit, and not proportional accuracy of interpolations at each scale. Substrate, which contains silt, sand, gravel, cobble, boulder, and land, will have r^2 values for all substrate sizes combined, as well as for each type. Breaking substrate into categories allows for a more complete view of predictive model accuracy loss across scales. Comparisons were made between predicted and baseline maps at each resolution for percent match. Following Sheehan and Welsh [12] root mean square error was also calculated for all depth, flow velocity, and substrate types in predicted maps.

Interpolated depth, flow velocity, and substrate values were subjected to ordinary least squares regression (OLS) in ArcMap to show proportional accuracy in comparison to the original 0.1 m² scale. Null, or -9999 values which occurred at site edge boundaries were not included in ordinary least squares calculations of r^2 . As previously stated, the baseline, or base, data scale for this study was 0.1 m². Values of r^2 resulting from scale regressions will allow for equivalent comparisons between scale because all scale interpolations were extracted and appended to the initial, and largest, data scale at all x,y coordinate locations for each site.

Further analysis included mapping of local OLS residuals at both Grayling and Wapiti creeks for each x,y coordinate. Mapping residuals is important because it allows for unique examination of regional accuracy of interpolations. Maps were created showing over and under estimation of each coordinate with standard deviation values classes ranging from -2.5 to 2.5. Residual maps were created by performing OLS regression on extracted natural neighbor interpolated values for all scales compared to base scale. Residual maps also serve to visually compliment proportional r^2 values and illustrate locations of inaccuracies.

Results

Results indicate depth, flow velocity, and substrate prediction accuracy for Wapiti and Grayling creeks decreased similarly as scale departed from the initially collected base data scale of 0.1 m² (Figures 2-6). Grayling Creek had consistently lower levels of RMSE than Wapiti Creek for flow, very similar RMSE values for depth, and nearly identical RMSE values at all scales for substrate (Figures 2-6). Depth and flow retained accuracy more effectively than substrate. At both sites, for all variables, predictive accuracy decay occurred in a log linear fashion as indicated by increasing RMSE as scale decreased from base scale (Figures 2-6). Natural neighbor interpolations maintained overall site integrity relatively well throughout all scales, though pattern decay had a visual smoothing effect on variable transitional boundaries and the site as a whole (Figures 7-8). As scale decreased from 0.1 m² to 13.38 m² smoothing of data resulted in loss of the full range of possible depths, flows, and more frequent atypically located substrate values. As indicated by r^2 values, interpolations made with data further removed from the initial scale they maintained overall sight integrity even as map accuracy continued to decay (Table 2).

Natural neighbor interpolated maps of Grayling Creek and Wapiti Creeks created from study scales provided visual confirmation of the smoothing process observed in RMSE, r^2 , and residual standard deviation locations and values (Figure 7,8). Map complexity decreased as scale of interpolation became more course for all habitat variables at both sites. Initial data loss of extreme depths, flow variation atypical of other values in the area, and atypically placed substrate were shown by the decrease in display complexity. Base scale maps contained a wide range of shape, occurrence and variability for each habitat variable. As scale decreased from base, the range, variability of shape, and site complexity gave way to smoother more

homogenous habitat values. During this smoothing, location and shape of deep areas, thalweg and flow zones, and substrate depositional areas were generally well maintained even to the terminal scale (Figure 7, 8).

Specific to Wapiti Creek

Depth, flow, and overall substrate prediction accuracy loss moving from fine to coarse scale degraded in a logarithmic fashion. This is demonstrated by larger RMSE values as scale moved away from 0.1 m² (Figures 2-4). Substrate, broken into categories, had overall RMSE values range from 0.6 to 1.30 (Figure 7). For individual substrate size ranges, sand maintained the lowest RMSE values, never exceeding 0.75 (Figure 7). Accuracy loss, which was notably different between habitat variables, is illustrated by linear regressions comparing different scales. The r^2 of regressions showed proportional match decreased for each habitat variable as scale moved away from base scale (Table 2). Flow models decreased the most dramatically, where r^2 values ranged from 0.86 to 0.34. Regressions also demonstrated models created from the coarsest scale, 148 times larger in terms of cell size, were able to outperform finer scales for some variables (Table 2).

Maps of depth, flow velocity, and substrate residuals displayed positive and negative prediction trends in the form of standard deviation at each coordinate location (Figure 7). Predictions closer to the initial large scale showed highly localized fluctuation in standard deviation values surrounding atypical habitat occurrence. Similar standard deviation fluctuations occurred at habitat transitional boundaries between dissimilar habitat types such as deep areas, or anomalous flow readings. Ordinary Least Squares regressions demonstrated that all models tended to underestimate deeper sections of river, and overestimate shallow sections, creating a

smoothing effect of both edge boundary and depth transition zones. This predictive smoothing effect increased in size proportional to the habitat feature as scale decreased.

An example of smoothing on Wapiti Creek occurs when examining depth where smoothing at scales smaller than 0.1 m^2 limited possible range of depths. By the smallest scale, the range of depths produced by interpolations was 0-44.9 cm, the furthest removed from base scale, rather than 0-83 cm and was the largest limitation. Though the upper flow limit was reduced by 38 cm, the mean shifted only from 22.8 to 19.7 centimeters by scale 12, further illustrating habitat smoothing effect. In addition, while the first quantile shifted lower towards zero, Median (26) and third quantile (34) remained the same across all scales. One of the problems with grid sampling is that it completely disregards habitat structure, and seeks to capture the full range of habitat values by correct selection of study scale. One potential method of compensating for value range loss would be reinsertion of the full value range into interpolations. A second method would be similar to that proposed by Sheehan and Welsh [12] in which data selection is non-random, and based on key habitat locations such as edge boundaries, extremes of values, and transitional zones.

Maps of interpolations and residuals showed spatial smoothing as scale departed from the base scale, but also showed ability to maintain overall site integrity even at scales far removed from the original. Smoothing effect is clearly illustrated by visual inspection, which was part of our methodology, and is a key factor in spatial analysis allowing for additional insight by illustrating statistical results (Figures 7, 8)[11, 13]. Data smoothing as resulting in the loss of detail in depth, flow, and substrate maps is relatively consistent (Table 2).

Specific to Grayling Creek

Prediction accuracy of depth, flow, and substrate on Grayling Creek degraded in a logarithmic fashion moving from 0.1 m² to 13.38 m² scale. Larger RMSE values for substrate occurred as scale decreased from 0.1 m² (Figures 2-6). Similar to Wapiti Creek, accuracy loss was different between habitat variables (Table 2). Substrate had r^2 values range from 0.86 to 0.54 (Table 2). Like Wapiti Creek, Grayling Creek linear regressions comparing base scale to scales 2-12 showed that percent (proportional) match decreased for each habitat variable as scale decreased and moved away from the base scale.

Grayling Creek maps of depth, flow velocity, and substrate residuals displayed similar positive and negative trends in standard deviation as Wapiti Creek (Figure 8). Like Wapiti Creek, Grayling Creek OLS regressions demonstrated the model tendency to underestimate deeper sections of river, and overestimate shallow sections as scale decreased and moved away from the original. The smoothing effect observed in Wapiti Creek was again shown in all three habitat variables of Grayling Creek (Figure 8). However, Grayling Creek, maintained a higher level of overall predictive accuracy for depth and flow by scale 12 than Wapiti Creek.

Discussion

Our study demonstrates that habitat data at a single scale (0.1 m² in this study) can be successfully used to accurately predict habitat at larger spatial scales. Two important big picture inferences can be drawn from this study. First, as data scale used to create predictive maps of stream habitat variables departs from the original resolution, model accuracy decay occurs in a log linear fashion. As accuracy decays, the differing scales retain predictive capability high enough to produce accurate and practical (functional, easily interpreted, informative) maps of stream habitat data. This is important because adequate accuracy retention between scales gives

the capability to provide multi-scale inferences from a single data set. Results of the study provided clues to the amount and type of accuracy decay for key habitat variables depth, flow velocity, and substrate in streams. Second, the predictive maps of depth, flow velocity, and substrate allow use of a single data scale for practical inferences at various spatial scales within similar reaches on streams of related order, classification, and geographic location. This capability imbeds more value of information into a single datasets. Increased value of information is integral to stream biologists and managers seeking guidance and insight on current conditions and lays the groundwork for future studies without necessitating additional expenditures of time and budget.

More specifically, our study documented a trend in data accuracy loss associated with natural neighbor interpolation of stream habitat data. The study revealed that sufficient accuracy is maintained for inferences at scales both larger and smaller than the original collection scale. While the data collected at the scale of 13.38 m² raster size is just as accurate as the scale of 0.1 m² in terms of each x,y coordinate, there appears to be a finite range of predictive effectiveness for each scale in terms of the value of information stored within each x,y coordinate. This follows the point made by Fisher and Rahel [13] that data collected at small scale may eliminate detail important to ecological studies found in larger scales. Following this logic, the more accurately data are maintained when interpolating at other scales, the more demonstrably useful a dataset.

In their review, Fisher and Rahel [13] indicated neither the location, mode, or speed in which accuracy was lost, nor the point in which accuracy of models was no longer acceptable for stream habitat assessment or other management purposes; this study directly addressed those questions. Our data indicated a structured accuracy loss of stream habitat variables depth, flow,

and substrate that can be described visually and statistically as maintaining usefulness at scales far removed from the initial collection scale. This is supported by RMSE, proportional r^2 values, and visual maps of sites showing initially high levels of model accuracy which fall off, but do not entirely diminish even at scales 11 times removed from the original 0.1 m². Further, our results also demonstrated stream habitat variables depth, flow, and substrate followed a concise linear relationship for type and amount of accuracy lost when moving away from the initially collected data scale. Each of these facts is integral to understanding predictive stream models and interpolations of aquatic habitat data and provides needed structure to the framework of stream habitat interpolation methodology.

Initial scale of collected data and stream size influence the range of scales at which the data set retains usefulness for predictive purposes. For instance Wapiti Creek showed a drop in predictive accuracy below 60% at the fifth scale removed (from the original) for all habitat variables. The Wapiti site encompassed a series of three pool/riffle zones, while Grayling Creek was a single pool riffle interface (three to four times the scale of Wapiti). Because of this stream scale difference, results may have indicated presence of a threshold for predictive accuracy purely associated with stream size. This makes ecological sense, in that a larger order stream may have proportionally larger habitat patches which in turn maintains any scale's predictive accuracy at further reaching scales.

Smoothing Effect

In addition to monitoring proportional accuracy retention of data between scales, the ability to quantify the type and location of accuracy loss is also important. By observing the combined results of RMSE trends and residual values as a guide to natural neighbor interpolative inaccuracies, it is possible to see a detailed progression of errors caused by departure from initial

scale when interpolating stream habitat variables. The general effect for all stream habitat variables was what can be termed a smoothing effect, which eliminated values on the extreme of depth and flow, and tended to under or over-estimate substrate transitional zones. As illustrated by Purkait [23] and mentioned by Knighton [26], substrate sorting often occurs in abrupt transitions zones, which was maintained through repeated smoothing by interpolations.

An additional facet of stream habitat modeling and assessment which is important for management and ecological studies is the ability to maintain spatial integrity of site boundaries and habitat transition zones. This allows the accurate measurement of area of available habitat. In streams, the amount and distribution of available habitat are closely tied to species occurrence and species diversity, and their examination aids in ecological study at varying spatial scales [2, 3, 27-30]. Maps of interpolations and regression residuals also aid in clarifying the scalability of stream habitat variables depth, flow velocity, and substrate by showing the specific locations of strong and weak predictions. Wapiti and Grayling creek residual maps showed specific site location smoothing by over estimating shallow areas, and underestimating deeper areas for depth. Flow values were similarly over and under-estimating. Substrate maps illustrated that transitional zones experienced the most smoothing where abrupt changes occurred, or where transitions included disparate categories like sand transitioning to boulder, rather than gravel or silt.

The ability to move between scales from a single data set has value to assessment, management, and analysis of stream habitat data because it lowers the costs associated with the collection of data. This is accomplished by allowing inferences at scales other than that which the data were originally collected. As pointed out by Rastettler et al. [31] it is possible to make inferences at coarse scale from fine scale data. This study expands that assertion, demonstrating

the reverse is also true. Many stream habitat and aquatic organism studies have since acknowledged the importance of analysis between scales for aquatic studies including benthic macro invertebrates, fish, and anthropogenic factors [3, 5, 6, 32-35]. In our study RMSE and r^2 values and ordinary least squares regression illustrated that when moving away from the collected data scale accuracy loss initially occurs quickly, but still allows for effective predictions even at scales far removed from the initial data collection scale.

This study demonstrated that it is possible to maintain a level of accuracy when moving from one spatial scale to another with stream habitat data. We have found that the accuracy of output maps can be maintained even when the scale is simplified from the original. However, an important aspect of accuracy of stream habitat models yet to be addressed is the matter of perspective and need; what level of accuracy is acceptable for a biological study? Data taken at coarse scale, such as catchment, may be 95% accurate at that scale, yet only allow for 40% accuracy of stream habitat when used to model stream habitat at the finer reach scale. In many predictive spatial models, accuracy levels approaching or above 70% or r^2 above 0.70 are considered successful. While we are not advocating any particular threshold for acceptable accuracy in this study, the ability to scale a single stream habitat data-set and still maintain accuracy far above the 0.70 r^2 level is a valuable tool for stream management and assessment purposes. In particular, management of lotic systems is based on best available, though often incomplete information. Therefore, providing more complete information to managers without increasing budget or data collection needs is an important product of this study.

By combining data from micro and macro scales, biologists are able to generate diverse conclusions about streams and aquatic organisms. Spatial analysis of stream habitat across spatial scales has been tapped for analysis needs specifically catered to address species

occurrence based on habitat connectivity and occurrence between scales [7, 27]. This study acknowledges that need by an initial demonstration of success in moving from large or small scale from a single data set. Inherent in that demonstration are increases in analysis potential and potential limitation of need for redundant data collection by acknowledging connectivity between scales. There are various methods for general stream assessment and classification including Rosgen, Montgomery-Buffington, and rapid visual habitat methods such as those used by state and government entities [36, 37]. Use of classification and assessments of stream habitat has become more specific and detailed over time in order to match management need and current knowledge base.

Acknowledgments

We thank Kyle Hartman, Stephen Kite, and Walt Kordek for input on the scientific and editing processes. We thank Richard Sheehan for data collection assistance and Nicole Ten Eyck for recording of data. Reference to trade names does not imply endorsement of commercial products by the U.S. government.

Literature Cited

1. Petty, J.T., et al. *Quantifying instream habitat in the upper Shavers Fork basin at multiple spatial scales*. in *Proceedings of the Annual Conference of the Southeastern Association of Fish and Wildlife Agencies*. 2001.
2. Petty, T. and G. Grossman, *Patch selection by mottled sculpin (Pisces: Cottidae) in a southern Appalachian stream*. *Freshwater Biology*, 1996. **35**(2): p. 261-276.

3. Thompson, A.R., J.T. Petty, and G.D. Grossman, *Multi-scale effects of resource patchiness on foraging behaviour and habitat use by longnose dace, *Rhinichthys cataractae**. *Freshwater Biology*, 2001. **46**(2): p. 145-160.
4. Valavanis, V.D., et al., *Modelling of essential fish habitat based on remote sensing, spatial analysis and GIS*. *Hydrobiologia*, 2008. **612**(1): p. 5-20.
5. Gido, K.B., et al. *Fish-habitat relations across spatial scales in prairie streams*. in *American Fisheries Society Symposium 48*. 2006. Bethesda, Maryland: American Fisheries Society.
6. Urban, D.L., *Modeling ecological processes across scales*. *Ecology*, 2005. **86**(8): p. 1996-2006.
7. Allan, J.D., D.L. Erickson, and J. Fay, *The influence of catchment land use on stream integrity across multiple spatial scales*. *Freshwater biology*. Oxford, 1997. **37**(1): p. 149-161.
8. Levin, S.A., *The problem of pattern and scale in ecology: The Robert H. MacArthur Award Lecture*. *Ecology*, 1992. **73**(6): p. 1943-1967.
9. Gotway, C.A. and L.J. Young, *Combining Incompatible Spatial Data*. *Journal of the American Statistical Association*, 2002. **97**(458): p. 632-648.
10. Cooper, S.D., et al., *Implications of scale for patterns and processes in stream ecology*. *Austral Ecology*, 1998. **23**(1): p. 27-40.
11. Gergel, S., et al., *What is the value of a good map? An example using high spatial resolution imagery to aid riparian restoration*. *Ecosystems*, 2007. **10**(5): p. 688-702.
12. Sheehan, K.R. and S.A. Welsh, *An interpolation method for stream habitat assessments*. *North American Journal of Fisheries Management*, 2009. **29**(1): p. 1-9.

13. Fisher, W.L. and F.J. Rahel, *Geographic information systems applications in stream and river fisheries*, in *Geographic information systems in fisheries.*, W.L. Fisher and F.J. Rahel, Editors. 2004, American Fisheries Society: Bethesda, MD. p. 49–84.
14. Childs, C., *Interpolating surfaces in ArcGIS spatial analyst*, in *ArcUser2004*, ESRI: California. p. 32-35.
15. Bennett, R.J., R.P. Haining, and D.A. Griffith, *Review article: The problem of missing data on spatial surfaces*. Annals of the Association of American Geographers, 1984. **74**(1): p. 138-156.
16. Kratzer, J.F., D.B. Hayes, and B.E. Thompson, *Methods for interpolating stream width, depth, and current velocity*. Ecological Modelling, 2006. **196**(1-2): p. 256-264.
17. Le, N.D., W. Sun, and J.V. Zidek, *Bayesian multivariate spatial interpolation with data missing by design*. Journal of the Royal Statistical Society. Series B (Methodological), 1997. **59**(2): p. 501-510.
18. MacKay, D.J.C., *Bayesian interpolation*. Neural Computation, 1992. **4**(3): p. 415-447.
19. Sambridge, M., J. Braun, and H. McQueen, *Geophysical parametrization and interpolation of irregular data using natural neighbours*. Geophysical Journal International, 1995. **122**(3): p. 837-857.
20. Zimmerman, D., et al., *An experimental comparison of ordinary and universal kriging and inverse distance weighting*. Mathematical Geology, 1999. **31**(4): p. 375-390.
21. Jopling, A.V. and D.L. Forbes, *Flume study of silt transportation and deposition*. Geografiska Annaler. Series A, Physical Geography, 1979. **61**(1/2): p. 67-85.
22. Lunt, I.A., J.S. Bridge, and R.S. Tye, *A Quantitative, Three-dimensional depositional model of gravelly braided rivers*. Sedimentology, 2004. **51**(3): p. 377-414.

23. Purkait, B., *Patterns of grain-size distribution in some point bars of the Usri River, India*. India Journal of Sedimentary Research, 2002. **72**(3): p. 367-375.
24. Smith, G.H.S. and R.I. Ferguson, *The gravel-sand transition along river channels*. Journal of Sediment Research, 1995. **65**(2a): p. 423-430.
25. Moran, P.A.P., *Notes on continuous stochastic phenomena*. Biometrika, 1950. **37**(1-2): p. 17.
26. Knighton, D., *Fluvial forms and processes: A new perspective* 1998, New York: Oxford University Press Inc.
27. Winemiller, K.O., A.S. Flecker, and D.J. Hoeinghaus, *Patch dynamics and environmental heterogeneity in lotic ecosystems*. Journal of the North American Benthological Society, 2010. **29**(1): p. 84-99.
28. Townsend, C.R., *The patch dynamics concept of stream community ecology*. Journal of the North American Benthological Society, 1989: p. 36-50.
29. Pringle, C.M., et al., *Patch dynamics in lotic systems: The stream as a mosaic*. Journal of the North American Benthological Society, 1988. **7**(4): p. 503-524.
30. Hanski, I. and M.E. Gilpin, *Metapopulation dynamics*. Nature, 1998. **396**(6706): p. 41-49.
31. Rastetter, E.B., et al., *Aggregating fine-scale ecological knowledge to model coarser-scale attributes of ecosystems*. Ecological Applications, 1992. **2**(1): p. 55-70.
32. Weigel, B.M., et al., *Relative influence of variables at multiple spatial scales on stream macroinvertebrates in the Northern Lakes and Forest ecoregion, USA*. Freshwater Biology, 2003. **48**(8): p. 1440-1461.

33. Wang, L., et al. *Effects of levels of human disturbance on the influence of catchment, riparian, and reach-scale factors on fish assemblages*. 2006. American Fisheries Society.
34. Roth, N.E., J.D. Allan, and D.L. Erickson, *Landscape influences on stream biotic integrity assessed at multiple spatial scales*. Landscape Ecology, 1996. **11**(3): p. 141-156.
35. Kotliar, N.B. and J.A. Wiens, *Multiple scales of patchiness and patch structure: A hierarchical framework for the study of heterogeneity*. Oikos, 1990. **59**(2): p. 253-260.
36. Rosgen, D.L., et al., *A stream classification system*, in *First North American Riparian Conference* 1985: Rocky Mountain Forest and Range Experiment Station. p. 91-95.
37. Montgomery, D.R. and J.M. Buffington, *Channel-reach morphology in mountain drainage basins*. Bulletin of the Geological Society of America, 1997. **109**(5): p. 596.

Table 1. Raster cell size in meters squared for each predictive scale.

Scale	Cell Size Area in Meters Squared
Base (1)	0.1
2	0.37
3	0.84
4	1.49
5	2.32
6	3.36
7	4.55
8	5.95
9	7.53
10	9.29
11	11.24
12	13.38

Table 2. Wapiti Creek base scale (0.3 m) compared to scales 2, 4, 6, 8, 10, and 12. Base scale values of substrate, 1 (silt), 2 (sand), 3 (gravel), 4 (cobble), 5 (boulder), and 6 (land) are plotted on the x axis and comparative scales 2, 4, 6, 8, 10, and 12 are plotted on the y axis. Plots were fit with a regression line and associated r^2 values to show proportional accuracy of scale predictions to the base scale.

Little Wapiti					
Substrate		Depth		Flow	
R-squared	Scale	R-squared	Scale	R-squared	Scale
0.90	2	0.88	2	0.86	2
0.80	4	0.73	4	0.74	4
0.74	6	0.59	6	0.67	6
0.57	8	0.25	8	0.37	8
0.51	10	0.4	10	0.34	10
0.53	12	0.35	12	0.41	12
Grayling					
Substrate		Depth		Flow	
R-squared	Scale	R-squared	Scale	R-squared	Scale
0.86	2	0.95	2	0.95	2
0.78	4	0.91	4	0.94	4
0.69	6	0.86	6	0.87	6
0.67	8	0.87	8	0.9	8
0.56	10	0.78	10	0.78	10
0.54	12	0.81	12	0.78	12

Figure 1. Example of x,y coordinate point grid for Grayling Creek showing 4,950 data locations each containing depth, flow, and dominant observed substrate (at intermediate axis) information. Natural neighbor interpolation of the points created the baseline visual map.

Figure 2. Root mean square error values for natural neighbor predicted maps of depth values on Grayling and Wapiti creeks. Scales 1-12 are found on table 1 and range from 0.1 to 13.38 m² cell size.

Figure 3. Root mean square error values for natural neighbor predicted maps of flow velocity values on Wapiti and Grayling creeks. Scales 1-12 are found on table 1 and range from 0.1 at the base scale to 13.38 m² cell size at scale 12.

Figure 4. Root mean square error values for predicted maps of all substrate values at Wapiti and Grayling Creeks. Scales 1-12 are found on table 1 and range from 0.1 to 13.38 m² cell size.

Figure 5. Wapiti Creek substrate RMSE values with a logarithmic trend line applied showing r^2 values for each. Increase in RMSE as scale moves away from the baseline reference scale shows a progressive tapering effect of accuracy loss as predictive scale moves away from the baseline. Sand substrate size maintained the smallest RMSE change between scales

Figure 6. Grayling Creek substrate RMSE values with a logarithmic trend line applied showing r^2 values for each. Increase in RMSE as scale moves away from the baseline reference scale shows a progressive tapering effect of accuracy loss as predictive scale moves away from the baseline.

Figure 7. Wapiti and Grayling creek residual maps showing depth, flow, and substrate standard deviations for each of the 3,630 x,y coordinate points at the site. Highly localized regions of standard deviation variation in scale one progress to larger regions of similar standard deviation values as scale becomes smaller, referred to in the text as a smoothing effect. Distribution pattern type of standard deviation types moves from random in scale 2 to clustered in scale 12.

Figure 8. Grayling Creek residual maps showing depth, flow, and substrate standard deviations for each of the 3,630 x,y coordinate points at the site. Abrupt localized changes in standard deviation are more prevalent in scale one. A more gradual change in standard deviation, or smoothing effect, may be seen in scale 12.

Chapter 4

Advantages of Geographically Weighted Regression for Modeling Benthic Substrate in Two Greater Yellowstone Ecosystem Streams

Abstract

Stream habitat assessments are commonplace in fish management, and often involve non-spatial analysis methods for quantifying or predicting habitat, such as ordinary least squares regression (OLS). Spatial relationships, however, often exist among stream habitat variables. For example, water depth, water velocity, and benthic substrate sizes within streams are often spatially correlated and may exhibit spatial non-stationarity. Thus, analysis methods should address spatial relationships within habitat datasets. In this study, OLS and a recently developed method, geographically weighted regression (GWR), were used to model benthic substrate from water depth and water velocity data at two stream sites within the Greater Yellowstone Ecosystem. For data collection, each site was represented by a grid of 0.1 m² cells, where actual values of water depth, water velocity, and benthic substrate class were measured for each cell. Accuracies of regressed substrate class data by OLS and GWR methods were calculated by comparing maps, parameter estimates, and r^2 . For analysis of data from both sites, AICc indicated the best approximating model for the data resulted from GWR and not from OLS. Adjusted r^2 values also supported GWR as a better approach than OLS for prediction of substrate. This study supports GWR (a spatial analysis approach) over non-spatial OLS methods for prediction of habitat for stream habitat assessments.

Introduction

Assessment of fish habitat is an integral part of fish management [1]. Analysis methods for fish habitat assessments, however, have not always addressed issues of spatial correlation inherent within habitat data. For example, stream assessments frequently focus on three spatially autocorrelated habitat variables; water depth, water velocity, and substrate sizes [2-5]. Spatial correlation within a data set can be useful for predictive purposes [6], because a data point, such as a habitat value, can be predicted from a proximately-located value [7]. Predictions of stream habitat are useful for aquatic habitat managers, in part, because of the potential to reduce field time and data collection, thus relieving budget and time constraints. Researchers that use prediction methods in aquatic habitat assessments, however, have often ignored or overlooked spatial correlation and non-stationarity of data through the use of non-spatial analyses, such as ordinary least squares (OLS) regression [8]. Spatial non-stationarity of benthic substrate, depth, and flow velocity refers to the tendency of these variables to vary and remain unfixed in the geographic space of the stream. Geographically weighted regression (GWR), a recently-developed method for analysis of spatial data, may be a useful analysis tool and a method to address spatial non-stationarity issues inherent in fish habitat assessments.

Geographically weighted regression is a modified regression method for the analysis of spatial data with inherent qualities of spatial auto-correlation and non-stationarity. More specifically, GWR was created to model data with heterogeneity, which stream habitat variables often exhibit [9, 10]. The method was first introduced in the mid 1990's [10], and later applied to ecological studies [11, 12], and oceanic fisheries research and management [13, 14]. Researchers have examined and compared the applicability of GWR for analysis of spatial data relative to

that of other regression methods [15-17]. However, such analyses have not extended far into the field of fisheries science (e.g. Windle, [14]).

Geographically weighted regression is often compared to OLS because it illustrates benefits of using a spatial non-stationarity approach in statistical models [9-11, 18]. The GWR method differs from that of OLS in several ways. The standard expression for OLS is $y_i = \beta_0 + \sum_k \beta_k x_{ik} + \varepsilon_i$ which is modified to include the expression of (u_i, v_i) , in both the β_0 and β_k facets of the GWR regression formula [10]. One may think of (u_i, v_i) , as an x,y coordinate. In effect, a coordinate location is added to each data observation or i and thus gives the new equation $y_i = \beta_0(u_i, v_i) + \sum_k \beta_k(u_i, v_i)x_{ik} + \varepsilon_i$ [9, 10], which is useful in the analysis of spatial data [19]. Further, GWR does not follow the assumption of homoskedacity, or static variance, but instead calculates a specific variance for data within a zone, or search radius of each predictor variable. Variance becomes dynamic and allows for a more accurate regression of data when non-stationarity (heteroskedacity) is present [10, 20]. Geographically weighted regression calculates specific variance for each coordinate point. In this way, GWR may be considered a local rather than a global model, because variance is not an averaged or single value. This is important for stream habitat models because variables such as substrate, depth and flow, often exhibit dynamic localized variation along habitat transition zones.

In part due to GIS, there has been dramatic increase in the body of knowledge of stream habitat in the past two decades, particularly regarding spatial qualities of stream habitat data. However, examination of some stream ecological data has become hampered due lack of understanding of issues such as spatial heterogeneity, non-stationarity, spatial autocorrelation, and multicollinearity. Such issues indicate need for better statistical methodologies, a need also acknowledged in the overlapping field of species occurrence and ecological study [12]. Due to

these issues, prior methods of analysis of fisheries data may no longer fully address needs of biologists working with temporally rapid change in stream habitat [21]. This has led to a decade long call to action for new methodology and focus within the fisheries community [21-24].

Geographically weighted regression, as detailed earlier, is an appropriate method for demonstrating the usefulness of spatially aware statistics when modeling stream habitat data. Therefore, this study was established to examine the hypothesis that GWR produces a better fit when regressing stream habitat data in comparison to OLS regression. The study focused on evaluating model fit for benthic substrate at stream sites using data on water depth and water velocity. To our knowledge, others have not reported on the use of GWR to produce regression fit specific to this type of stream habitat data. This study is relevant and important to fish habitat assessments and aquatic habitat scientists because it emphasizes the need to consider often ignored and overlooked spatial qualities within stream habitat data. The study also evaluates an analysis method that could potentially improve data predictions through use of spatial data relationships. Further, because many ecological studies do not address issues of autocorrelation or non-stationarity, yet are using statistics on data which violate their assumptions, this study would clarify ability of geographically weighted regression to produce accurate results when run in such a condition.

Methods

Study Sites

The first of our two study sites was located in the Gallatin National Forest on Little Wapiti Creek, Montana, (111°16'53"W, 45°2'20"N). The second study site was located in another tract of Gallatin National Forest on Grayling Creek, Montana (111°6'16"W, 44°48'16"N). Sampling

occurred between July 18, 2008 through August 4, 2008 during daylight hours and periods of stream flow at seasonally normal rates post snow melt. The study sites were selected due to relatively remote and undisturbed locations in the Greater Yellowstone Ecosystem, a mountainous region including characteristic precipitation regime of heavy winter snows and dry summers. Monthly precipitation averages 4.47 cm for both July and August, and a yearly average of 72.11 cm. The two sites were 28.2 km apart. Little Wapiti creek feeds the Gallatin River watershed, and Grayling Creek drains directly into Hebgen Lake. Free range cattle use the drainage containing Little Wapiti watershed during summer, though the site was undisturbed at the time of sampling; there was no apparent stream bed alteration at either site due to cattle or other disturbances.

Sites were delineated by grid cells (0.3 m resolution per cell, an area of 0.1 m²) using a fifty meter tape measure, laser rangefinder, and flagging (later removed). The Little Wapiti Creek site measured 33.5 meters long by a maximum of 10 meters wide (Figure 1). The Grayling Creek site measured 27.5 meters long by 20.1 meters wide (Figure 1). Stream depth (cm, top-setting wading rod) and mean water velocity (m/s at 60% depth, Marsh-McBirney Flowmate 2000) were measured at the center of each cell. Substrate was measured on a continuous scale in millimeters diameter at intermediate axis diameter from 0mm to >300 mm for each cell. Thus, actual, continuous values of water depth, water velocity, and substrate size were recorded for each 0.1 m² cell from the grid of each study site. Two data sets (one for each site) were created using ArcMap 10 (ESRI, 2003) and Microsoft Excel 2010. In ArcMap, corner points for each study site were georeferenced and exported to Microsoft Excel. Next, x,y coordinates were calculated for the remainder of cells in the site grid and appended to the initial dataset of water depth, water

velocity, and substrate size. The final datasets were imported back to ArcMap 10 for statistical analyses.

For all OLS and GWR regressions a two variable model was used for the prediction of substrate size; depth and flow velocity were explanatory variables modeling substrate size, which was the dependent variable for all models. For each dataset, OLS regression (ArcToolbox, ArcMap 10) provided model residuals for Moran's I, Jarque-Bera and Joint Wald statistics, the Variance Inflation Factor (VIF), and the Koenker (BP) studentized tests. Moran's I was used to test for significant levels ($p < 0.05$) of spatial autocorrelation in regression residuals [25, 26]. The Jarque-Bera statistic tests for significant levels of normalcy ($p < 0.05$) in the distribution of residuals of OLS regressions [27]. The Joint Wald statistic tests for model significance ($p < 0.05$) [28]. The VIF value is an explanatory variable redundancy check (multicollinearity, $p < 0.05$). The Koenker (BP) studentized test was used to test for significant levels ($p < 0.05$) of non-stationarity in OLS residuals (Koenker 1981), as well as examine heteroscedasticity of data for all models on both Grayling and Wapiti sites. According to Brunsdon et al. (2005), significant levels of spatial autocorrelation and non-stationarity of residuals indicate that the data are better suited for GWR analysis. The geographically weighted regression tool located in ArcToolbox (ArcMap 10) was used for this step.

GWR regressions were further structured in the following manner: For the Grayling site, a 20 neighbor search radius was chosen for GWR regressions and for Little Wapiti a 23 neighbor search radius was chosen. Search radius was selected for each site by starting at eight neighbors (the smallest possible) and increasing by one until the model was stable and had no misspecified values in the resulting attribute table. Akaike's information criterion (AICc) was used to determine the best approximating model to the data, and allowed for comparison between

modeling approaches (OLS vs. GWR). Both OLS and GWR analysis methods allowed for estimation of AICc and associated model selection statistics following standard methods [29]. The full suite of AIC values including delta AIC, and model likelihood, were back calculated from regression summary tables.

Maps of GWR parameter estimates of standardized residuals, condition number, local r^2 , and predicted benthic substrate were created for comparison of OLS and GWR. Ordinary least squares maps of predicted benthic substrate values and standardized residuals were also created for comparison to GWR. Local r^2 indicates model performance in specific locations, and shows where the model experiences difficulty. In conjunction with condition number, which demonstrates local multicollinearity, the two parameters may indicate locations of weak and strong model performance.

Results

Ordinary least squares indicated both flow and depth were significant, non-redundant variables in the model (Table 1). Significant values for robust t for depth and flow supported their significance in the model and that neither should be removed from the model (Table 1). The Koenker (BP) statistic was significant for both sites indicating spatial non-stationarity (Table 1). The significant Koenker (BP) statistic led us to examine the Joint Wald statistic, which indicated that our model was significant (Table 1).

Observations of heteroskedacity were supported by significant results from Moran's I of OLS residuals which found significant levels ($p < 0.05$) of spatial auto-correlation among water depth, flow, and substrate size (Tables 1, 2). Significant levels of spatial autocorrelation and significant non-stationarity of OLS residuals provides strong support for using GWR analysis

instead of OLS. The OLS VIF for both sites showed lack of multicollinearity among explanatory variables depth and flow (Table 2, [10, 30, 31]).

For both Little Wapiti and Grayling sites, adjusted r^2 values and AICc values supported GWR as a better modeling approach for predicting benthic substrate data than that of OLS regression (Tables 2, 3). For Wapiti Creek, adjusted r^2 values from OLS were considerably lower than those from GWR (Table 2). Similarly, for Grayling Creek, adjusted r^2 values from OLS were lower than those from GWR (Table 2). Also, for the analyses of both Little Wapiti and Grayling creeks, lower AICc values supported models from the GWR analysis, but did not support models from OLS (Table 3).

Geographically weighted regression captures the natural spatial patterns of stream depth, flow velocity, and substrate more effectively than OLS regression (Figures 2,3,4). GWR captures regional variation in stream habitat variables depth, velocity, and substrate which is integral to an accurate prediction by incorporating the spatial non-stationarity found in the variables[10, 31]. This is demonstrated by examining GWR predicted values, which visually match the benthic substrate deposition pattern more closely than OLS regressions (Figures 2,3). It is also shown by observation of GWR standardized residuals which tend to keep within 0.5 standard deviations more frequently than OLS. Moran's I of residuals for GWR regressions also trends towards random, which indicates the GWR model better accounted for the spatial non-stationarity found at Little Wapiti and Grayling than OLS. Geographically weighted regression local r^2 and condition number maps were effective in illustrating multicollinearity which was not effectively captured by OLS regressions (Figure 4, Table 1). Some condition numbers from GWR indicated potential multicollinearity at localized regions of both sites which was not noticed by OLS diagnostics (Tables 1, 2, Figure 4).

Discussion

As mentioned in the introduction, this study originated as response to the specific need to quantify stream habitat data models using spatial regression in the form of GWR rather than the non-spatial alternative of OLS. Recent inclusion of spatial analysis in fisheries research and management afforded by GIS has exposed potential for gains in understanding fisheries processes and their management [32-36]. A growing body of fisheries literature addresses the importance of analysis of spatial qualities of habitat and habitat patches across stream scales [34]. Increased efforts within the fisheries science community towards spatial analysis has only served to cement its usefulness in the field. This study directly addressed the issue of regression model fit for spatially non-stationary stream habitat data of depth, flow, and substrate when using GWR versus OLS regression.

It may be stated from this study that geographically weighted regression has advantages for modeling benthic substrate in streams. We recommend geographically weighted regression for analysis of stream habitat variables, in part, because it addresses spatial variation, auto-correlation, and non-stationarity issues [10, 31]. Auto-correlation of stream habitat such as substrate is well documented and indicate Moran's I values were accurately reflecting both this phenomenon and the inability of OLS regression to accurately represent it. Known variance generally increases accuracy of predictions [37]. We are supporting this statement by adding detail to our knowledge of variance by addressing spatial non-stationarity, heteroskedacity, and auto-correlation within the data set. The Koenker (BP) statistic demonstrated spatial non-stationarity with a significant p-value and further cemented the rationale for moving to GWR analysis over non-spatial OLS for stream habitat assessments such as benthic substrate [10, 38].

Our study addressed another specific concern about ecological variables exhibiting spatial auto-correlation, non-stationarity, and multicollinearity by showing that accurate predictive regressions may occur when using GWR when such conditions are present. Spatial auto-correlation and non-stationarity are difficult if not impossible to remove from some ecological studies and analysis. Nonetheless, regression is often used in ecological studies though such conditions violate regression assumptions and may cause less accurate or cloudy results [8, 39]. Further, that OLS regression did not statistically notice multicollinearity, which is clearly present in localized maps created by GWR, illustrates the risks of using static variance in the presence of spatial non-stationarity. Our results support this assertion by demonstrating the lack of accuracy in OLS regression which was remedied when using GWR. Ability to identify localized areas of multicollinearity provides opportunity to further tune predictive models. Because many ecological studies exhibit spatial non-stationarity, fisheries biologists and managers using regression statistics on stream habitat data should be aware of the potential loss in regression fit if those conditions are not addressed with methods such as GWR.

In some ecological models, autocorrelation and multicollinearity may cause overfit of model variables, and thus inflate perceived accuracy. This may be illustrated by the following example: Flow influences substrate deposition just as substrate influences flow. Similarly, a submerged boulder in a stream will alter the flow of water around it, before it, after it, and thus influence depth and surrounding substrate by its flow alterations. The same may be said of all points and variables within the stream, each connected by the bond of location. While over-estimation may exist to some extent, the addition of spatial location of data in GWR regression in this study clearly shows improved model fit which would be useful in analysis of stream

habitat data. By addressing the spatial autocorrelation and non-stationarity within stream habitat variables, GWR has kept intact one of the most important qualities of stream habitat.

Following the intent of Brunson, Fotheringham, and Charlton, the goal of GWR analysis specific to stream habitat variables is more accurate prediction of existing or future habitat for management and ecological study. Accurate prediction of habitat allows for better assessment of current stream condition and helps guide management practices for better stream quality in the future [1]. Stream ecological studies often focus on habitat parameters because of their close tie to species occurrence and number [12, 40]. Studies of fish life history and behavior often use habitat such as depth, velocity, and substrate as foundation level variables; increasing accuracy of habitat assessment and prediction would aid in these endeavors [2, 41-43]. Geographically weighted regression accomplishes increased regression accuracy of stream habitat variables depth and velocity to predict substrate.

Looking to necessary future changes, stream assessments established in recent decades fail to incorporate detailed spatial qualities into protocols. For example, river morphology and classification are frequently assessed using methods such as Rosgen or Montgomery-Buffington [44-46]. Though rough spatial qualities are initially measured, specific spatial attributes of habitat patches such as edge boundaries, area, perimeter, and distance to other habitat are not part of those assessment products except to locate the study site on a map. This creates an analysis gap between current knowledge of connectivity between stream networks and the data collected to analyze them. The evolving landscape of statistical tools available for spatial analysis across scales in all aspects of fishery science requires analysis- including habitat assessment, just as it does in species occurrence [12].

To sum, this study demonstrated the importance of maintaining spatial relationships in stream habitat data by GWR in comparison to OLS regression. Shape and dimension and location are spatial qualities of habitat patches like substrate which otherwise would not be accurately portrayed by OLS regression, but are highly important in stream and ecological studies [41-43, 47]. The study further demonstrated the important role of non-stationarity and spatial auto-correlation to modeling stream habitat data; geographically weighted regression produces better model fit than OLS regression under these circumstances. Though spatial autocorrelation of stream habitat variables is frequently passed over in standard regression analysis such as OLS, the advantages of GWR in producing unambiguous, accurate predictions emphasizes the need to consider spatial relationships within stream habitat data. The use of GWR for stream habitat modeling has important implications for data collection, habitat assessment, and habitat prediction because of its ability to provide more accurate results than non-spatial regression analysis.

Literature Cited

1. Ian, M., *The importance of physical habitat assessment for evaluating river health*. Fresh Water Biology, 1999. **41**(2): p. 373-391.
2. Aadland, L.P., *Stream habitat types: Their fish assemblages and relationship to flow*. North American Journal of Fisheries Management, 1993. **13**(4): p. 790-806.
3. Lazorchak, J.M., et al., *Environmental monitoring and assessment program surface waters: Field operations and methods for measuring the ecological condition of non-wadeable rivers and streams* 2000: US Environmental Protection Agency, Office of Research and Development.
4. Lazorchak, J.M., et al., *Field operations and methods for measuring the ecological condition of wadeable streams*, E.P. Agency, Editor 1998, US EPA. p. 270.
5. Sheppard, J.D. and J.H. Johnson, *Probability-of-use for depth, velocity, and substrate by subyearling coho salmon and steelhead in Lake Ontario tributary streams*. North American Journal of Fisheries Management, 1985. **5**(2b): p. 277-282.
6. Legendre, P., *Spatial autocorrelation: trouble or new paradigm?* Ecology, 1993. **74**(6): p. 1659-1673.

7. Sheehan, K.R. and S.A. Welsh, *An interpolation method for stream habitat assessments*. North American Journal of Fisheries Management, 2009. **29**(1): p. 1-9.
8. González-Megías, A., J.M. Gómez, and F. Sánchez-Piñero, *Consequences of spatial autocorrelation for the analysis of metapopulation dynamics*. Ecology, 2005. **86**(12): p. 3264-3271.
9. Charlton, M., S. Fotheringham, and C. Brunsdon, *Geographically weighted regression*. NCRM Methods Review Papers, 2005(NCRM/006).
10. Fotheringham, A.S., M. Charlton, and C. Brunsdon, *The geography of parameter space: an investigation of spatial non-stationarity*. International Journal of Geographical Information Science, 1996. **10**(5): p. 605-627.
11. Kupfer, J. and C. Farris, *Incorporating spatial non-stationarity of regression coefficients into predictive vegetation models*. Landscape Ecology, 2007. **22**(6): p. 837-852.
12. Austin, M., *Species distribution models and ecological theory: a critical assessment and some possible new approaches*. Ecological Modelling, 2007. **200**(1-2): p. 1-19.
13. Wang, L., et al. *Effects of levels of human disturbance on the influence of catchment, riparian, and reach-scale factors on fish assemblages*. 2006. American Fisheries Society.
14. Windle, M.J.S., et al., *Exploring spatial non-stationarity of fisheries survey data using geographically weighted regression (GWR): an example from the Northwest Atlantic*. ICES Journal of Marine Science, 2009. **67**(1): p. 145-154.
15. Ali, K., M.D. Partridge, and M.R. Olfert, *Can geographically weighted regressions improve regional analysis and policy making?* International Regional Science Review, 2007. **30**(3): p. 300.
16. Gao, X., Y. Asami, and C.J.F. Chung, *An empirical evaluation of spatial regression models*. Computers and Geosciences, 2006. **32**(8): p. 1040-1051.
17. LeSage, J.P., *A family of geographically weighted regression models*, in *Advances in Spatial Econometrics*, L. Anselin, R.J. Florax, and S.J. Rey, Editors. 2001, Springer: Berlin, Germany. p. 241-263.
18. Lo, C.P., *Population estimation using geographically weighted regression*. GIScience & Remote Sensing, 2008. **45**(2): p. 131-148.
19. Mennis, J., *Mapping the results of geographically weighted regression*. The Cartographic Journal, 2006. **43**(2): p. 171-179.
20. Zimmerman, D., *Geographically weighted regression*. Journal of the American Statistical Association, 2003. **98**(463): p. 765-767.
21. Poole, G.C., C.A. Frissell, and S.C. Ralph, *In-stream habitat unit classification: Inadequacies for monitoring and some consequences for management*. Journal of the American Water Resources Association, 1997. **33**(4): p. 879-896.
22. Caddy, J.F., *Fisheries management in the twenty-first century: will new paradigms apply?* Reviews in Fish Biology and Fisheries, 1999. **9**(1): p. 1-43.
23. Minns, C.K., *Science for freshwater fish habitat management in Canada: current status and future prospects*. Aquatic Ecosystem Health & Management, 2001. **4**(4): p. 423-436.
24. Stephenson, R.L. and D.E. Lane, *Fisheries management science: A plea for conceptual change*. Canadian Journal of Fisheries and Aquatic Sciences, 1995. **52**(9): p. 2051-5056.
25. Moran, P.A.P., *Notes on continuous stochastic phenomena*. Biometrika, 1950. **37**(1-2): p. 17.
26. Moran, P.A.P., *The statistical analysis of the Canadian lynx cycle. I. Structure and prediction*. Australian Journal of Zoology, 1953. **1**(1): p. 163-173.

27. Jarque, C.M. and A.K. Bera, *Efficient tests for normality, homoscedasticity and serial independence of regression residuals*. Economics Letters, 1980. **6**(3): p. 255-259.
28. Agresti, A., *Categorical data analysis*. Vol. 359. 2002: John Wiley and Sons.
29. Burnham, K.P. and D.R. Anderson, *Model selection and multimodel inference: a practical information-theoretic approach* 2002: Springer Verlag.
30. Fotheringham, S., C. Brundson, and M. Charlton, *Geographically Weighted Regression & Associated Techniques* 2002, Chichester, England: Wiley.
31. Charlton, M., S. Fotheringham, and C. Brunsdon, *Geographically weighted regression*. ESRC National Centre for Research Methods NCRM Methods Review Papers NCRM/006, 2005. **NCRM/006**.
32. Meaden, G.J., *Application of GIS to fisheries management*. Marine and Coastal Geographical information Systems, 2000: p. 205–226.
33. Meaden, G.J. *GIS in fisheries science: foundations for the new millennium*. in *First International Symposium on GIS in Fishery Science*. 2001. Saitama, Japan: Fisheries Research Group.
34. Winemiller, K.O., A.S. Flecker, and D.J. Hoeinghaus, *Patch dynamics and environmental heterogeneity in lotic ecosystems*. Journal of the North American Benthological Society, 2010. **29**(1): p. 84-99.
35. Fisher, W.L., *Future of geographic information systems in fisheries*, in *Geographic Information Systems in Fisheries*, W.L. Fisher and F.J. Rahel, Editors. 2004, American Fisheries Society: Bethesda, MD. p. 259-266.
36. Fisher, W.L. and F.J. Rahel, *Geographic information systems applications in stream and river fisheries*, in *Geographic information systems in fisheries*., W.L. Fisher and F.J. Rahel, Editors. 2004, American Fisheries Society: Bethesda, MD. p. 49–84.
37. Robertson, G.P., *Geostatistics in ecology; Interpolating with known variance*. Ecology (Durham), 1987. **68**(3): p. 744-748.
38. Koenker, R., *A note on studentizing a test for heteroscedasticity*. Journal of Econometrics, 1981. **17**(1): p. 107-112.
39. Graham, M.H., *Confronting Multicollinearity in Ecological Multiple Regression*. Ecology, 2003. **84**(11): p. 2809-2815.
40. Rosenfeld, J., *Assessing the habitat requirements of stream fishes: an overview and evaluation of different approaches*. Transactions of the American Fisheries Society, 2003. **132**(5): p. 953-968.
41. Beisel, J.N., et al., *Stream community structure in relation to spatial variation: the influence of mesohabitat characteristics*. Hydrobiologia, 1998. **389**(1): p. 73-88.
42. Gorman, O.T. and J.R. Karr, *Habitat structure and stream fish communities*. Ecology, 1978. **59**(3): p. 507-515.
43. Gido, K.B., et al. *Fish-habitat relations across spatial scales in prairie streams*. in *American Fisheries Society Symposium 48*. 2006. Bethesda, Maryland: American Fisheries Society.
44. Rosgen, D.L., *A classification of natural rivers*. Catena, 1994. **22**(3): p. 169-199.
45. Rosgen, D.L., et al., *A stream classification system*, in *First North American Riparian Conference* 1985: Rocky Mountain Forest and Range Experiment Station. p. 91-95.
46. Montgomery, D.R. and J.M. Buffington, *Channel-reach morphology in mountain drainage basins*. Bulletin of the Geological Society of America, 1997. **109**(5): p. 596.

47. Wintle, B.A. and D.C. Bards, *Modeling Species-habitat Relationships with Spatially Autocorrelated Observation Data*. Ecological Applications, 2006. **16**(5): p. 1945-1958.

Table 1. Ordinary least squares diagnostics indicating the model variables are neither redundant nor exhibit multicollinearity (VIF), are statistically significant (Joint Wald), exhibits non-stationarity (Koenker (BP)), and exhibits residuals distributed non-normally (Jarque-bera).

OLS Diagnostics	VIF	Koenker (BP) P-value	Robust T	Robust T P-value	Jarque-bera Statistic P-value	Joint Wald Statistic P- value
Little Wapiti	-	0.04*	-	-	0.00*	0.00*
Intercept	-	-	14.76	0.00*	-	-
Flow	1.09	-	18.46	0.00*	-	-
Depth	1.09	-	12.18	0.00*	-	-
Grayling		0.00*	-	-	0.00*	0.00*
Intercept	-	-	7.29	0.00*	-	-
Flow	2.73	-	8.93	0.00*	-	-
Depth	2.73	-	5.85	0.00*	-	-

**significant at the 0.05 level*

Table 2. Ordinary least squares regression and GWR r^2 and adjusted r^2 values for depth and flow velocity relationships to substrate on Little Wapiti and Grayling creeks, Montana. Akaike's information criterion values are also included. Depth and flow velocity are explanatory variables, and substrate is the predicted value for all regressions.

Location	Adjusted		Moran's I of			Model
	R-squared	R-squared	Residuals	Z-Score	P-value	
Little Wapiti	0.62	0.79	-0.02	-0.94	0.35	GWR
	0.25	0.25	0.37	21.20	0.0*	OLS
Grayling	0.68	0.78	0.03	2.07	0.039*	GWR
	0.28	0.28	0.51	33.12	0.0*	OLS

Table 3. Akaike information criterion parameters from GWR and OLS regressions for Little Wapiti and Grayling creek sites. Depth and flow velocity are the explanatory variables for all models. Substrate is the predicted value for all regressions.

Location	Regression Type	Search Radius	N	K	Log Likelihood	AIC	AICc Value	Delta AIC	Model Likelihood	Akaike Weight
Little Wapiti	GWR	20 neighbor	1848	4	-10064.91	20137.81	20137.83	0	1	1
	OLS	-	1848	4	-10598.48	21204.95	21204.97	1067.14	0	0
Grayling	GWR	14 neighbor	4288	4	-11113.09	22234.19	22234.2	0	1	1
	OLS	-	4288	4	-24387.75	48783.51	48783.52	26549.32	0	0

Figure 1. Grayling Creek wetted width at upstream boundary is 17.7m, and 20.1m at the downstream boundary; maximum width is 20.1m and minimum width is 13.9m. Little Wapiti Creek wetted width at upstream boundary is 3.7m, and is 2.3m at the downstream boundary; maximum width is 9.4m and minimum width is 1.8m.

Figure 2. Grayling creek standardized residual maps and predicted benthic substrate from GWR and OLS regressions.

Figure 3. Standardized residuals and predicted benthic substrate maps from GWR and OLS regressions.

Figure 4. Condition number and local r^2 maps of Little Wapiti and Grayling creeks created by GWR.

Heterogeneity and Site Detail

Grayling Creek

GWR
Standardized Residuals

OLS
Standardized Residuals

GWR
Predicted Values

Standardized Residuals (a, b)

- < -2.5 Std. Dev.
- -2.5 - -1.5 Std. Dev.
- -1.5 - -0.5 Std. Dev.
- -0.5 - 0.5 Std. Dev.
- 0.5 - 1.5 Std. Dev.
- 1.5 - 2.5 Std. Dev.
- > 2.5 Std. Dev.

Substrate Size in mm (c,d,e)

- Land
- 0.01 - 0.50
- 0.51 - 1.00
- 1.01 - 39.00
- 39.01 - 143.00
- 143.01 - 300.00

OLS
Predicted Values

Actual Values

Little Wapiti Creek

Local R-squared and Condition Number

Chapter 5

Geographically Weighted Regression Models and Role of Moran's I in Evaluating Heterogeneous Benthic Substrate in Streams

Abstract

Stream benthic substrate is frequently assessed due to its role in stream ecology. Recent efforts have focused on assessing its spatial complexity. However, spatial heterogeneity can limit such efforts by obscuring valuable relationships within the stream and reducing accuracy of their statistical analysis. Statistics such as geographically weighted regression and Moran's I have been developed to measure and account for spatial heterogeneity. Within stream studies, the use of geographically weighted regression is increasing; therefore it is important to know the effects of benthic substrate spatial heterogeneity on such regressions. Though initially created as a scale indicating dispersed, random, or clustered spatial distribution, Moran's Index has been suggested as a measure of spatial heterogeneity. Using geographically weighted regression and Moran's I as an indicator of spatial heterogeneity, this study examined substrate modeling on 40 quadrats on four streams in two separate geographic regions over 1,583 m². Results indicated accuracy of geographically weighted regressions of substrate decline as Moran's I values decreased from one to zero. By establishing a link between geographically weighted regression results (accuracy of regressions in the form of r^2) and benthic substrate heterogeneity present on a site, this study demonstrated the impact and importance of addressing the condition within a stream dataset. This relationship had a 0.66 r^2 value over all study sites. The correlation implies amount of spatial heterogeneity must be addressed as an independent variable in benthic substrate habitat

assessments. Stream studies containing spatial benthic substrate data which fail to do so may result in error propagation or incorrect interpretation of stream habitat data and assessments.

Introduction

Within stream studies, spatial analysis of streams and the role of heterogeneity is experiencing intense growth (Cooper et al. 1997; Levins 1969; Palmer et al. 1997; Winemiller et al. 2010). More specifically, creation of maps and analysis of spatial data in streams are becoming a staple product in government assessment regimes because of its importance to physical habitat assessment, management, and ability to provide more cost effective study structure (Bickers 2003; Gergel et al. 2007; Meaden 2001). While stream heterogeneity provides challenges for stream analysis and assessment, it is also an issue impeding accurate map creation and spatial variable analysis (Zhang and Gove 2005). These problems exist in part because heterogeneity has the potential to limit accuracy of interpolations and regressions. Accuracy limitations pass along product impacts to maps, their assessment, and analysis (Zhang and Gove 2005). Heterogeneity tends to cause excessive over and under estimation of habitat variables which limits map and prediction accuracies (Zhang and Gove 2005). Therefore, spatial heterogeneity in streams requires additional study to assess the impact it has on accuracy of mapping and assessment of stream habitat variables.

Previous studies have quantified the occurrence of spatial heterogeneity in streams (Cooper et al. 1997). However, these studies often address *functional heterogeneity* surrounding organism occurrence and not with *structural heterogeneity* directly. The studies and methodology examined tend to deal with the same topic (functional heterogeneity) and neither discuss heterogeneity's role from a standpoint of effective maps nor quantify levels of spatial

heterogeneity. Spatial heterogeneity studies have been limited due to lack of tools (Cooper et al. 1997), however newly available statistical tools in the form of spatially aware geographically weighted regression have alleviated that issue (Fotheringham et al. 1996).

Measurement of spatial distribution of data typically includes Moran's I, which may be used as a measure of heterogeneity (Anselin 2002; Anselin et al. 1996; Zhang and Gove 2005). Heterogeneity is also made up of two main components, those being complexity and variability (Li and Reynolds 1995). In addition, a study may examine structural heterogeneity and functional heterogeneity (Li and Reynolds 1995). Structural heterogeneity focuses entirely on the structure of habitat, and not its effects on ecological function. Moran's I measures structural heterogeneity by looking at spatial distribution of data within a study area. Moran's I values range from -1, which indicate a fully dispersed spatial pattern, to 1 which indicates a clustered, ordered spatial pattern. A zero value indicates a random spatial pattern (Anselin et al. 1996; Moran 1953).

Results from geographically weighted regression such as r^2 values, if closely correlated to Moran's I, would demonstrate a direct effect on their accuracy by heterogeneity. Geographically weighted regression r^2 values provide a measure of proportional model prediction accuracy. It is reasonable to posit that combining information from GWR analysis and Moran's I of benthic substrate would allow direct application of more efficient study and identification of valuable (spatial) stream habitat data for purposes of assessment and study. To our knowledge, this type of analysis combination has not been performed on stream habitat data.

Our hypothesis states that efforts to model benthic substrate in streams become less accurate as spatial heterogeneity increases. In addition, this study will demonstrate if model accuracy may be aided by applying Moran's I and regression residual analysis if a correlation is

found between regression results and Moran's I values. This may be shown by using geographically weighted regression and Moran's I values from 40 sub-sites on four stream reaches containing varying levels of heterogeneity. If the null model is not accepted, model accuracy (r^2) for each sub site would decrease as heterogeneity increases, and would be observed by correlation between Moran's I and regression r^2 . In this way, we will quantify stream habitat heterogeneity and show its impact on prediction and modeling of stream habitat variables. Demonstration of this occurrence would be highly useful to aquatic habitat managers because it would quantify accuracy loss due to spatial heterogeneity in what would otherwise be a model with high r^2 . In addition, knowledge of accuracy loss due to spatial heterogeneity allows the condition to be addressed and accounted for by adjustments in sampling regime and data analysis.

Methods

Study Sites and Data Collection

Field data were collected from four wadeable streams for this study. Two streams were located in the Greater Yellowstone Ecosystem, Gallatin National Forest and two were located in West Virginia; one each in Monongalia and Kanawha Counties. The first of our two western sites was located in the Gallatin National Forest on Little Wapiti Creek, Montana, (111°16'53"W, 45°2'20"N). The Wapiti Creek site measured 33.5 meters long by 10 meters wide (Figure 1). The second western study site was located in another tract of Gallatin National Forest on Grayling Creek, Montana (111°6'16"W, 44°48'16"N; Figure 1). The Grayling Creek site measured 27.5 meters long by 18 meters wide. The Elk River site was located downstream of the effluent of Big Sandy Creek in Kanawha County, West Virginia (81°21'3.857"W,

38°29'20.73"N) and measured 22 meters wide, by 27 meters long. The second eastern study site was located on Aarons Creek, which lies within the Monongahela River system in Monongalia County, West Virginia (79°56'0.625"W 39°37'8.69"N).

Sites were delineated by grid cells (0.3 m resolution per cell, an area of 0.1 m²) using a fifty meter tape measure, laser rangefinder, and flagging (later removed). Stream depth (cm, top-setting wading rod) and mean water velocity (m/s at 60% depth, Marsh-McBirney Flowmate 2000) were measured at the center of each cell. Substrate was measured on a continuous scale in millimeters diameter at intermediate axis diameter from 0 mm to >300 mm for each cell. Thus, actual, continuous values of water depth, water velocity, and substrate size were recorded for each 0.1 m² cell from the grid of each study site. Two data sets (one for each site) were created using ArcMap 10 (ESRI, 2010) and Microsoft Excel 2010. In ArcMap, corner points for each study site were georeferenced and exported to Microsoft Excel. Next, x,y coordinates were calculated for the remainder of cells in the site grid and appended to the initial dataset of water depth, water velocity, and substrate size. The final datasets were imported back to ArcMap 10 for statistical analyses.

Analysis Methods and data collection

Using ArcMap 10, 10 random points with 1.5 meter circular buffers were generated within each study site, termed quadrats. Circular buffers were chosen to avoid directional bias caused by substrate depositional pattern and depth in the stream. Coordinate points from depth and substrate data layers were then clipped to the randomly selected quadrats. Random quadrats with no variation in substrate were not subjected to analysis, and a new random point and buffer was selected within the site to replace it and create a new quadrat. Geographically weighted regression needs some heterogeneity to run the model otherwise a null or misspecified value is

returned. For instance, if a random point had only sand, and did not vary in depth, there would not be enough variation within the site to run the GWR statistic, resulting in a null or misspecified value for that area. In this manner, a total of 40 random quadrats with no misspecification were generated.

Moran's I (Global) values were calculated from observed substrate values found in each of the 40 random subsamples using substrate size as the analysis layer. Moran's I values were recorded for each site (Table 1). A simple GWR with a 30 neighbor adaptive search radius was conducted for each layer using depth as variables to explain substrate. Values of r^2 were recorded for each site (Table 1). To demonstrate the change in regression accuracy due to heterogeneity, the series of 40 random quadrats were sorted and plotted in descending order of r^2 . Moran's I was then plotted and a line was fit to Moran's I, demonstrating accuracy of the trend in comparison to r^2 . Plotting Moran's I and geographically weighted regression r^2 values in this way demonstrate directly the role of heterogeneity on accuracy of geographically weighted regressions.

Results

As Moran's I values of benthic substrate decrease from 1 to 0 GWR prediction values also decrease, as indicated by an r^2 of 0.66 when depth was used to model benthic substrate (Figure 1). Moran's I run on substrate for all 40 sub-sites ranged from 0.93 to 0.16 (Table 1). Results of GWR regressions were stable; there were no model warnings of severe design problems, overly high condition numbers, or regions of miss-specification of any one feature (x,y coordinate) in the models generated by the regression.

There were no negative Moran's I values generated; All values ranged between random (0) and clustered (1). All Moran's I values above 0.30 were significant within a 95% confidence interval, indicating that the clustered pattern observed in sub sites was not likely due to chance alone. Moran's I values close to zero (those below 0.30) were not significant, as one would expect when indicating random distribution of points. Geographically weighted regressions run with substrate as dependent variable and depth as explanatory variable resulted in r^2 values ranging from 0.979 to 0.247 (Table 1). Agreement of the slope of the line applied to Moran's I as geographically weighted regression r^2 values descended was 0.66 (Figure 1).

Discussion

When modeling benthic substrate, there is a correlation between heterogeneity and accuracy of GWR regressions. Results indicate when spatial heterogeneity of stream habitat data moves from a clustered distribution to a random pattern that GWR prediction of benthic substrate accuracy decreases. Evidence of this is visible in the strong correlation between r^2 values and Moran's I values when plotted across all sites. Our results also show increasing heterogeneity obscures analysis of spatial relationships among stream habitat variables; depth is known to be well correlated to substrate and stream benthic structure yet regressions in the presence of random spatial distribution were highly inaccurate. Thus spatial heterogeneity was shown to be directly responsible for limiting the accuracy of geographically weighted regression of stream habitat variables.

Results are strengthened by consistency between and among sites of varying size and geographic location. This is important because it indicates heterogeneity's influence is not limited to a particular geographic region or stream size. Implications of heterogeneity for stream

habitat modeling may not be understated because even potentially strong relationships between stream habitat variables could be overlooked due to low r^2 values in the presence of spatial heterogeneity. In turn this would render stream habitat study, analysis, and predictive maps less accurate and thus lead to improper conclusions on the part of managers attempting to interpret an incorrect result.

Implications for stream modeling

We recommend incorporating the quantification of spatial heterogeneity into habitat studies because of its impacts on study results (and thus conclusions) such as inflated mean square error (Dutilleul 1993) or decrease in GWR r^2 values, as shown in this study. There must be some measure of order and structure in a system for it to be modeled and predicted; a completely random system would have no basis for prediction as there would be zero correlation within and among variables. Based on our results, as substrate deposition becomes more entropic (random distribution and Moran's I closer to zero), GWR regressions lost ability to model the spatial pattern. Therefore, spatial heterogeneity decreases accuracy of stream habitat models and thus leads directly to inaccurate maps, assessments, and misguided decisions based on results made from spatially heterogeneous benthic substrate data. Therefore, accounting for the accuracy loss due to spatial heterogeneity in data has impacts on data requirements. This study demonstrates that impact directly.

Our study also indicates a stream reach with highly variable heterogeneous habitat would require more data to accurately measure and capture the shape, type, and location of each habitat patch than would one which was completely homogeneous. This is illustrated by the example that a square patch of 100% sand would only need points at each corner delineating the patch's shape, area, and substrate while a heterogeneous square patch of sand, gravel, cobble, and

boulder would require many more points in order to capture the site complexity. In this way, it can be stated that spatial heterogeneity causes increased data requirements in environmental studies, because r^2 values decline as heterogeneity increases, as shown by the results in Figure 1.

Naturally occurring complexities of stream habitat data aside, the smaller the data requirement to complete a task for stream spatial analysis, the smaller the total project cost (time, budget). Because our results showed GWR accuracy is directly impacted by spatial heterogeneity, accounting for the quality in a stream habitat assessment or analysis would create budget and time savings by allowing aquatic habitat scientists to vary the amount of data they needed to adequately capture the site. Following this logic, viewing heterogeneity as an independent stream variable becomes a viable approach because its measurement can provide a more accurate valuation of amount and type of data necessary for accurate stream habitat modeling, analysis, and assessment.

Specific to stream habitat variables depth and substrate, identifying locations of higher spatial value would have specific impacts on accuracy for purposes of modeling and prediction. However, selecting spatially valuable stream habitat data from the generic mass of that available at any given site may be a daunting task. Few entities have the time and money to completely catalogue entire streams at a high resolution. Identifying or assigning more valuable spatial data in streams to guide future research or aid in management endeavors would be an important product. Towards this end, results of this study show there are more valuable regions of spatial data such as zones of heterogeneity which tend to obfuscate variable relationships and would therefore require more data to model them effectively. This would be particularly important with interpolation procedures, another branch of predictive modeling used for map creation and analysis of stream habitat data. It is important because identification of zones of heterogeneity

would allow a modeler or scientist to know that more data are necessary to model that region, otherwise valuable information may be lost across scales. Our model results concur with terrestrial and broad method GWR mapping studies, in which similar behavior has been noted (Farber and Páez 2007; Mauricio Bini et al. 2009; Wang et al. 2005).

To sum, our results show it is possible to use Moran's I analysis in conjunction with GWR regression values to quantify heterogeneity's impact on spatial models and measure the amount of spatial heterogeneity on a stream reach. Further, our results provide meaningful guidance about amount of accuracy lost during GWR analysis to heterogeneity alone. This has implications for stream habitat modeling and assessments because measurement of heterogeneity would allow additional guidance of type, location, and amount of data necessary for collection by identifying regions prone to high levels of heterogeneity. In this way, a stream's benthic substrate may be assigned value based on location, an issue we are examining in studies of stream data selection for spatial models. More importantly, this study provides evidence that model fit is influenced by stream habitat heterogeneity. This fact allows appropriate adjustments in sampling protocol and analysis to offset those accuracy losses, or at the very least allow for realization that results may not be as accurate as desired, and for what reason.

Acknowledgments

We thank volunteers Richard Sheehan for data collection and Nicole Ten Eyck for data entry. Reference to trade names does not imply endorsement of commercial products by the U.S. government.

Literature Cited

- Anselin, L. 2002. Under the hood issues in the specification and interpretation of spatial regression models. *Agricultural Economics* 27(3):247-267.
- Anselin, L., A. K. Bera, R. Florax, and M. J. Yoon. 1996. Simple diagnostic tests for spatial dependence. *Regional Science and Urban Economics* 26(1):77-104.
- Bickers, A. N. 2003. Cost effective marine habitat mapping from small vessels using GIS, sidescan sonar and video. Coastal GIS. Australian Online Coastal Information, Geoscience Australia.
- Cooper, S. D., L. Barmuta, O. Sarnelle, K. Kratz, and S. Diehl. 1997. Quantifying spatial heterogeneity in streams. *Journal of the North American Benthological Society* 16(1):174-188.
- Dutilleul, P. 1993. Spatial heterogeneity and the design of ecological field experiments. *Ecology* 74(6):1646-1658.
- Farber, S., and A. Páez. 2007. A systematic investigation of cross-validation in GWR model estimation: empirical analysis and Monte Carlo simulations. *Journal of Geographical Systems* 9(4):371-396.
- Fotheringham, A. S., M. Charlton, and C. Brunsdon. 1996. The geography of parameter space: an investigation of spatial non-stationarity. *International Journal of Geographical Information Science* 10(5):605-627.
- Gergel, S., Y. Stange, N. Coops, K. Johansen, and K. Kirby. 2007. What is the value of a good map? An example using high spatial resolution imagery to aid riparian restoration. *Ecosystems* 10(5):688-702.

- Levins, R. 1969. Some demographic and genetic consequences of environmental heterogeneity for biological control. *Foundation papers in landscape ecology* 15(3):237-240.
- Li, H., and J. F. Reynolds. 1995. On Definition and Quantification of Heterogeneity. *Oikos* 73(2):280-284.
- Mauricio Bini, L., and coauthors. 2009. Coefficient shifts in geographical ecology: an empirical evaluation of spatial and non-spatial regression. *Ecography* 32(2):193-204.
- Meaden, G. J. 2001. GIS in fisheries science: foundations for the new millennium. Pages 3-29 *in* T. Nishida, P. J. Kailola, and C. E. Hollingsworth, editors. First International Symposium on GIS in Fishery Science. Fisheries Research Group, Saitama, Japan.
- Moran, P. A. P. 1953. The statistical analysis of the Canadian lynx cycle. I. Structure and prediction. *Australian Journal of Zoology* 1(1):163-173.
- Palmer, M. A., C. C. Hakenkamp, and K. Nelson-Baker. 1997. Ecological Heterogeneity in Streams: Why Variance Matters. *Journal of the North American Benthological Society* 16(1):189-202.
- Wang, Q., J. Ni, and J. Tenhunen. 2005. Application of a geographically weighted regression analysis to estimate net primary production of Chinese forest ecosystems. *Global Ecology and Biogeography* 14(4):379-393.
- Winemiller, K. O., A. S. Flecker, and D. J. Hoeinghaus. 2010. Patch dynamics and environmental heterogeneity in lotic ecosystems. *Journal of the North American Benthological Society* 29(1):84-99.
- Zhang, L., and J. H. Gove. 2005. Spatial assessment of model errors from four regression techniques. *Forest Science* 51(4):334-346.

Table 1. Raw geographically weighted regression r^2 values in descending order and labeled by site. Clendenin maintained both the highest and lowest r^2 values. Range of Moran's I and r^2 was well distributed across all sites. Moran's I index is of benthic substrate pattern.

Site	GWR R-Squared	Moran's Index of Substrate
Clendenin 0	0.979	0.852
Wapiti 5	0.967	0.929
Wapiti 3	0.965	0.880
Clendenin 5	0.956	0.815
Wapiti 6	0.932	0.875
Clendenin 8	0.922	0.677
Wapiti 8	0.881	0.609
Wapiti 0	0.880	0.661
Wapiti 2	0.851	0.677
Wapiti 7	0.843	0.535
Clendenin 2	0.808	0.547
Wapiti 1	0.777	0.691
Aaron's 2	0.774	0.764
Wapiti 4	0.770	0.414
Grayling 0	0.754	0.643
Aaron's 4	0.710	0.704
Aaron's 8	0.708	0.706
Aaron's 6	0.700	0.679
Clendenin 1	0.682	0.771
Grayling 6	0.666	0.668
Aaron's 7	0.641	0.746
Clendenin 7	0.639	0.680
Aaron's 5	0.636	0.615
Grayling 3	0.623	0.385
Aaron's 9	0.610	0.366
Grayling 7	0.524	0.274
Aaron's 1	0.518	0.399
Grayling 9	0.514	0.292
Aaron's 0	0.487	0.354
Grayling 1	0.442	0.332
Clendenin 6	0.440	0.419
Wapiti 9	0.430	0.244
Aaron's 3	0.418	0.174
Grayling 2	0.401	0.202
Grayling 5	0.393	0.572
Grayling 8	0.354	0.266
Clendenin 4	0.324	0.414
Clendenin 3	0.267	0.312
Grayling 4	0.249	0.332
Clendenin 9	0.247	0.379

Figure 1. The linear trend of Moran's I values decrease as geographically weighted regression r^2 values decrease for 40 stream sub-sites. This relationship shows regression values of substrate are correlated to Moran's I and thus spatial distribution type.

Chapter 6

Use of Geographically Weighted Regression to Model Benthic Substrate; A Comparison Between Large and Small Wadeable Streams

Abstract

Aquatic habitat assessments encompass large and small wadeable streams which vary from many meters wide to ephemeral. Differences in stream sizes within or across watersheds, however, may lead to incompatibility of data at varying spatial scales. Specifically, issues caused by moving between scales on large and small streams are not typically addressed by many forms of statistical analysis, making the comparison of large (>30m wetted width) and small stream (<10m wetted width) habitat assessments difficult. Geographically weighted regression (GWR) may provide avenues for efficiency and needed insight into stream habitat data by addressing issues caused by moving between scales. This study examined the ability of GWR to consistently model stream substrate on both large and small wadeable streams at an equivalent resolution. We performed GWR on two groups of 60 randomly selected substrate patches from large and small streams and used depth measurements to model substrate. Our large and small stream substrate models responded equally well to GWR. Results showed no statistically significant difference between GWR r^2 values of large and small stream streams. Results also provided a much needed method for comparison of large and small wadeable streams. Our results have merit for aquatic resource managers because they demonstrate ability to spatially model and compare substrate on large and small streams. Using depth to guide substrate modeling by geographically weighted regression has a variety of applications which may help manage, monitor stream health, and interpret substrate change over time.

Introduction

Wadeable stream habitat is monitored and studied across stream size, from ephemeral to many meters wide, to manage for various aspects of stream ecology including fish population dynamics and species occurrence (Gido et al. 2006; Rosenfeld 2003; Winemiller et al. 2010). Because of ties between habitat and population dynamics, wadeable stream assessment and monitoring protocols focus on quantifying key abiotic variables such as substrate and depth. Their assessments are used to create maps, monitor change, and categorize streams based on the information from those assessments. Differences in stream sizes within or across watersheds, however, may lead to incompatibility of data at varying spatial scales (Gotway and Young 2002; Scott et al. 2002). Issues caused by moving between scales on large and small wadeable streams are not typically addressed by many forms of statistical analysis, making the comparison of large (>30m wetted width) and small stream (<10m wetted width) habitat assessments difficult. Geographically weighted regression (GWR) is a new spatial modeling technique which may address issues of scale compatibility for important variables in stream habitat models.

Benthic substrate is a key variable in aquatic habitat assessments because of its biotic and abiotic importance. Therefore, efforts to maximize effectiveness of substrate assessment are important because substrate data collection is an integral activity which guides management of those streams (Hansen 2001; Ian 1999). Water depth is also an important and frequently measured stream habitat variable, and is closely correlated with sizes of benthic substrate (Allan and Castillo 2007; Gorman and Karr 1978; Knighton 1998). Water depth is measured within aquatic habitat assessments because it drives a variety of ecologic processes within the stream including location, food abundance, predator prey relationships, fish size, and reproductive success (Bradford and Heinonen 2008; Chisholm et al. 1987 Harvey and Stewart 1991).

However, connections among fish and habitat management protocols and current science often lag (Minns 2001). To close that gap, analysis and modeling of stream habitat have begun to incorporate spatial qualities of stream variables like benthic substrate and depth. Incorporating spatial qualities at an intricate level allows researchers and managers an avenue to gain insight and implement more effective management plans. Efficiency and accuracy of assessment are needed because scientists and managers must respond to human caused landscape and ecosystem wide alteration which are taking place at an ever increasing pace (Tilman et al. 2001; Vitousek et al. 1997). Because of the spatial correlation between depth and benthic substrate, depth can be a useful variable for predictive modeling of benthic substrate.

To monitor and assess substrate on large and small streams, multiple data analysis methods are often used to effectively capture and convey information across stream size. However, due to structural differences associated with changes in stream size and incompatibility of data at varying spatial scales, modeling techniques successful in predicting substrate on a small stream a few meters wide may not have application for a larger lotic body (Gotway and Young 2002; Rastetter et al. 1992; Wheatley and Johnson 2009). Therefore, determination of whether or not a single type of spatial modeling is capable of accurately mapping substrate on wadeable streams of both large and small size is a worthwhile undertaking. Further, such a determination would provide an appropriate basis for comparison of streams of varying size within and among studies if a method were successful in doing so.

As streams move through their course from headwaters to their terminus, their biotic and abiotic characteristics also change. As described in the river continuum concept, streams have a specific tiered structure at different stages along their path (Vannote et al. 1980). Stream habitat complexity typically increases with stream size due to increases in the types, amounts, and

organization of available habitat. Differences may include substrate depositional pattern, flow regime, oxygenation, turbulence, water clarity, and other abiotic factors. There are unique accuracy issues which may occur when attempting to assess and model streams because of these differences. Some relationships of habitat, such as depth changes, habitat spatial heterogeneity, and substrate depositional pattern do not always follow an ordered structure in relationship to stream size and may even be inverse to stream size. For instance, variation in depth may lessen as a stream becomes larger and closer to its terminus, substrate changes become less frequent, and flow velocity may become consistent because of a low gradient near a stream's terminus.

Ecological data interacts spatially with the environment and it is important to address those qualities or remove them from the dataset (Borcard et al. 1992). The importance of depth and benthic substrate to streams is closely tied to their location in the stream, and removal of spatial information would limit analysis and conclusion. Therefore, to accurately model stream habitat variables such as substrate and depth, the spatial qualities in stream data must be addressed within the analysis framework. For example, spatial non-stationarity, auto-correlation, and multi-collinearity are qualities of spatial data which may obscure meaningful results of statistical analysis such as regressions, one of the most commonly used types of analysis of stream habitat data (González-Megías et al. 2005; Legendre 1993). In addition to the pitfalls caused by spatial auto-correlation and multi-collinearity, there are issues caused by moving between scale though cross scale analysis is important for ecological analysis (Gotway and Young 2002; Levin 1992; Townsend et al. 2003; Urban 2005).

Geographically weighted regression has specific qualities which may adequately address data modeling issues caused by moving between scales. Geographically weighted regression is a modeling method which has been successfully used for analysis in fisheries studies by

incorporating spatial attributes of data including non-stationarity and spatial autocorrelation (Windle et al. 2009). The method addresses spatial non-stationarity by removing constraints of a global model and allowing for local variance to be calculated at each data location. Though still a type of linear regression, it allows for the often unaddressed variable of location to be addressed directly within the dataset. Spatial statistical methodologies such as GWR provide an appropriate framework to address whether or not they are able to perform equally as well on large or small wadeable streams. Geographically weighted regression may provide avenues for efficiency and needed insight into stream habitat such as substrate and depth data as well. More specifically, GWR was created to model data with heterogeneity, which stream habitat variables often exhibit (Charlton et al. 2005). The method was first introduced in the mid 1990's (Brunsdon 1995; Charlton et al. 2005; Fotheringham et al. 1996), and later applied to ecological studies (Austin 2007; Kupfer and Farris 2007), and oceanic fisheries research and management (Wang et al. 2001; Windle et al. 2009). Researchers have examined and compared the applicability of GWR for analysis of spatial data relative to that of other regression methods (Ali et al. 2007; Gao et al. 2006; LeSage 2001; Zhang and Gove 2005). However, such analyses have not extended far into the field of fisheries science (e.g. Windle 2009).

Geographically weighted regression appears to have the qualities to allow it to analyze stream variables like benthic substrate regardless of size. Therefore, the goal of this study is to examine the ability of GWR to model benthic substrate using depth values from both large and small wadeable streams at an equivalent level of accuracy. This would provide valuable streamlining to lotic spatial modeling and assessment by allowing a single method to be used regardless of stream size. Further, establishment of a single type of modeling statistic to

effectively compare both large and small streams would add clarity to the entire wadeable stream assessment process.

Methods

Study Sites

Depth and substrate were collected from four streams for this study. Two streams were located in the Greater Yellowstone Ecosystem, Gallatin National Forest bordering the western edge of Yellowstone National Park and two were located in West Virginia; one each in Monongalia and Kanawha Counties. A combined total of 17,040 x,y coordinate points, for a sum total of 1583 m² was recorded; each coordinate point represented a 0.3m x 0.3m (0.1 m²) cell on the stream site. Depth and dominant substrate size per quadrat (mm at intermediate axis diameter) were recorded individually at all coordinate points, which represented the centroid of each cell. Coordinate locations for each site were taken at the upstream left corner. Sites were chosen because they contained at least one pool and riffle interface.

The first of the two western sites was located in the Gallatin National Forest on Little Wapiti Creek, Montana, (111°16'53"W, 45°2'20"N). The second study site was located in another tract of Gallatin National Forest on Grayling Creek, Montana (111°6'16"W, 44°48'16"N). Sampling occurred between July 18, 2008 through August 4, 2008 during daylight hours and periods of normal stream flow. The Wapiti Creek site measured 33.5 meters long by 10 meters wide, for a total of 3,630 (0.093 m) cells, each representing depth, and dominant substrate type. The Grayling Creek site measured 27.5 meters long by 18 meters wide, for a total of 4,950 cells (0.093 m), each representing depth and dominant substrate size. The study sites were selected due to relatively remote, though easily accessible and undisturbed locations in the

Greater Yellowstone Ecosystem, a mountainous region including characteristic precipitation regime of heavy winter snows and dry summers. The two sites were 28.2 km apart. Little Wapiti Creek feeds the Gallatin River watershed, and Grayling Creek drains directly into Hebgen Lake. Free range cattle use the Little Wapiti watershed during summer, though both sites were relatively undisturbed and stream bed alteration was not attributable to cattle or other unnatural disturbances at the time of sampling.

Clendenin Shoals on Elk River is just downstream of the effluent of Big Sandy Creek in Kanawha County, West Virginia (81°21'3"W, 38°29'20"N). The modeled site measured 22 meters wide, by 27 meters long and contained 6,192 (0.093 x 0.093 m) cells. Unique among the four study sites, this site was located within the town limits of Clendenin. Housing and moderate urban development occur along the banks of the river, though flow and river dynamics are relatively undisturbed due to lack of urban, suburban, or farming development upstream of the study site. The second eastern study site was located on Aarons Creek, which lies within the Monongahela River system in Monongalia County, West Virginia (79°56'0"W 39°37'8"N). The site measured 24.33 x 8.66 m and contained 2,268 (0.093 m²) cells. There is sparse to moderate urban and suburban development along approximately 70% of this 13.5-km stream. The riparian area of the stream (5 – 50 m wide) is a mixture of pasture and mixed hardwood forest.

Data Collection

Study sites on all four water bodies were delineated by grid cells (0.3 m resolution per cell) using a fifty meter tape measure, laser rangefinder, and flagging (later removed). Habitat variables depth and dominant substrate were measured at the centroid of each cell along a secured tape measure crossing the entire site horizontally. This was repeated until the site

(minimum of one pool riffle interface) was captured in a complete grid of x,y coordinate points with habitat variables at each point.

Stream depth (cm, top-setting wading rod) was measured at the center of each cell. The dominant substrate size was also recorded in each cell along a continuous scale in millimeters from 0.05 to 300 mm based on intermediate axis diameter. Thus, actual values of water depth and substrate size were recorded for each 0.1 m² cell for each study site. Four data sets (one for each site) were created electronically using ArcMap 10 (ESRI, 2010) and Microsoft Excel 2010. In ArcMap, corner points for each study site were geo-referenced and exported to Microsoft Excel. Next, x,y coordinates were calculated for the remainder of cells in the site grid and appended to the initial dataset of water depth and substrate size. The final datasets were imported back to ArcMap 10 for analysis.

Prior to the use of geographically weighted regression, Ordinary Least Squares (OLS) regression was run on each site to establish appropriate need for geographically weighted regression (Brunsdon et al. 2000; Charlton et al. 2005). Output from OLS regression, the Koenker-BP test statistic (Koenker 1981), was examined to establish spatial non-stationarity variance significance. Moran's I was also run on all sites to indicate spatial pattern of the data (Moran 1953). If spatial autocorrelation was discovered in Moran's I values and residual maps of OLS regression indicated broad patterns of over and under estimation of values, then geographically weighted regressions were run.

Once appropriate use of GWR was established at each study site, GWR was run on all sites in entirety. In addition, 30 random points were generated within the stream boundaries at each site, resulting in a total of 120 random sample points (Figures 1, 2). Each of the 120 points had a buffer of 2.5 meters applied to it creating a sampling quadrat, and all points captured within that

quadrat were used for GWR regressions. Geographically weighted regression on Aarons Creek, Elk River, Grayling Creek, and Little Wapiti Creek sites produced 30 r^2 values each (one for each sampling quadrat). Thus, 60 r^2 values were recorded for two categories of streams; large (Elk and Grayling sites) and small (Little Wapiti and Aarons creek sites).

Additional details concerning GWR tests were as follow: substrate size was set as the dependent variable and depth was set as the explanatory variable. In order to create a standard comparable result across sites, regressions were run on each site using the same kernel type and bandwidth method. Each regression was observed to assure that no model misspecification occurred for any features. Specifically, GWR parameters for each pool riffle complex were run using an adaptive kernel type and bandwidth parameter.

Search radius was explored to provide maximum r^2 values and no model misspecification. Decreasing the search radius from the default (30) yielded higher overall r^2 values due to substrate depositional pattern which often occurs in highly localized regions and transitions abruptly. After exploring search radius, it was set to eight points for Grayling, Little Wapiti, and Aarons Creek and 16 points for Elk River, the largest site. For reference, values of r^2 were recorded for each site as a whole.

Results would provide two data sets for analysis; a population of 60 r^2 variables from large streams and a population of 60 r^2 variables from small streams. Variance between the data sets was examined using an f-test. Results from the f-test would determine the appropriate type of t-test. Comparison between large and small streams was accomplished by performing Welch's t-test assuming unequal variance on the two populations under the null hypothesis that the means of the two populations were not significantly different at an α -level of 0.01.

Results

Results from OLS for all sites had significant p-values for the Koenker-BP test statistic, indicating that non-stationary variance has made standard error of the regressions unreliable. Moran's I tests for all sites demonstrated a clustered pattern in the residuals from OLS regression; all Moran's I values were significant at an α level of 0.01, indicating there was less than a 1% chance this pattern was due to chance, and likely due to spatial non-stationarity qualities in the data. Per described methodology, GWR was run on all sites in entirety, and on all random samples of both large and small streams after appropriate use of the GWR statistic was established. Visual inspection of OLS regressions shows large areas of over and under estimation in prediction values of substrate along heterogeneous areas, and zones of substrate transition (Figures 1, 2).

Geographically weighted regression r^2 values of full study sites were 0.754 on Elk River, 0.839 on Aarons Creek, 0.871 on Grayling Creek, and 0.912 on Little Wapiti Creek. Random sampling and regression at 60 larger order and 60 smaller order stream quadrats produced r^2 values ranging from 0.170 (a highly heterogeneous quadrat) to 0.998 on Elk River, 0.698 to 0.930 on Aarons Creek, 0.699 to 0.990 on Grayling Creek, and 0.701 to 0.932 on Little Wapiti Creek. Test for variance indicated the variance of the two populations was not equal with a p-value < 0.01 . Means for r^2 of random quadrats were 0.83 for small streams and 0.79 for large streams. Results from the Welch's t-test demonstrated differences in means from substrate models of random samples of large and small streams were not significant as shown by a p-value of 0.22 at α level of 0.01.

Visual inspection of residual maps of GWR regressions indicated the statistic responded similarly to substrate for all streams, regardless of size, by minimizing over and under estimation

of predicted values of substrate. Minimization of zones of extreme standard deviation can be directly observed in GWR regressions by noticing decrease in locations returning standard deviation errors in the -1.5 to -0.5 and the 0.5 to 1.5 categories in comparison to those from OLS regressions on all streams (Figures 3, 4). The majority of predicted values that fell into those two categories were adjusted into the -0.5 to 0.5 standard error range, showing a marked improvement in overall site models for substrate in terms of standard error. Overestimation and underestimation of values was limited to highly heterogeneous localized regions in GWR regressions (Figures 3, 4).

Discussion

Our results establish the effectiveness of GWR as a modeling tool for wadeable streams of disparate size when analysis of spatial benthic substrate data is required. Specifically, our results show substrate is modeled in an equally effective manner on large or small wadeable streams when using GWR. This is supported by comparison of two large, and two small wadeable streams and values of r^2 from their geographically weighted regressions. Our results have implications to fisheries biologists and managers wishing to provide consistent, comparable assessment, analysis, and mapping results of stream habitat variables when their studies span multiple stream sizes. Implications include using a single regression type (GWR) to model stream habitat variables, potential for more reliable management of stream habitat when substrate is a key decision variable, and use of visual residual maps to gain insight into the results of substrate modeling.

The ability to use a single regression model to address spatial mapping and assessment of both large and small wadeable streams demonstrates geographically weighted regression's utility

to stream assessments and thus to fisheries science. It accomplishes this by minimizing the amount of statistical procedures necessary to properly map stream habitat variables such as substrate; the statistical analysis type used does not need to change in order to analyze and compare streams of varying size.

Our study indicates GWR's strength as a statistical procedure for modeling frequently assessed stream habitat spatial variables such as substrate. This is important in part because failure to address spatial qualities in a dataset when performing regressions may create ambiguous or erroneous results due to spatial autocorrelation (González-Megías et al. 2005; Legendre 1993). Using geographically weighted regression to model stream habitat data with acknowledged spatial qualities directly addresses the issues posed by spatial auto-correlation. It performs this task by removing a single global variance value used to calculate individual predictions within the model with local variance which is calculated using a specified search radius or parameter (Brunsdon 1995; Charlton et al. 2005). This change allows for more accurate representation of structure found within data exhibiting spatial non-stationarity (spatial variance), such as the depositional pattern of substrate. The usefulness of GWR to fisheries science and management can be seen by spatial variance often found in stream habitat variables; there are many frequently assessed variables found within a stream with spatial non-stationarity including flow velocity, temperature, location of large woody debris, and seasonal location of fish.

When modeling variables relevant to stream management such as substrate the ability of an analysis method to produce consistent results is particularly important. Consistent, unambiguous results lend themselves towards use in comparison because of confidence in the methodology and the added complexity of comparing results from different methods being removed. As stream size varies, our study indicates mean values of r^2 are consistent and do not

differ statistically when an appropriate GWR model is applied. Without consistent results the conclusions and management decisions drawn from them would be suspect, and at the very least, somewhat weaker.

Results of the study are in line with prior literature discussing fluvial geomorphologic relationships between depth and substrate. In our results the correlation between depth and substrate depositional pattern is shown by relatively high r^2 values between substrate and depth. We provide evidence that substrate may be effectively modeled using depth as the independent, or influencing variable when using geographically weighted regression to model wadeable stream habitat data. This statement is supported by accuracy levels of geographically weighted regression models of substrate on small and large wadeable streams are not significantly different as indicated by GWR r^2 values and t-test results.

Practical maps (maps which maintain both utility towards the named purpose and ease of use) are important aspects of interpreting results of stream habitat assessments, but are not often provided by statistical analysis as they are by GWR. Maps can be a valuable tool and provide insight which non-visual results would not provide (Gergel et al. 2007). Visual analysis of habitat type, amount, and fish populations on small streams does successful precedent (Hankin and Reeves 1988). In addition to practical maps of habitat data already produced by mapping stream habitat data in a geographic information system, geographically weighted regression provides maps in the form of residuals and predicted values for each coordinate location in the dataset. Therefore, visual interpretation of data and results is a useful benefit to the decision making process provided by geographically weighted regression when used in a program such as ArcMap 10.

An integral visual step in the GWR process for evaluating substrate in streams is establishment of the appropriate use of spatial regression rather than traditional ordinary least squares regression. Ordinary least squares regression is always run to demonstrate non-stationarity in the data, without which GWR would not be an appropriate course of action (Brunsdon 1995; Charlton et al. 2005). A visual byproduct of OLS regression is the residual map created to show amount of standard deviation of each predicted value at each data coordinate. Besides showing non-stationarity, substrate was clearly over and underestimated in a greater proportion of the study area than with GWR residual maps on all four main study sites. In this instance, visual comparison of GWR and OLS regression residual maps helped to illustrate the benefits of applied local variance for spatial stream data such as substrate, rather than a single variance. Specific to the comparison of large and small streams, visual inspection of OLS and GWR residual maps show that GWR reacts similarly to both large and small streams.

As seen in maps of large streams Elk River and Grayling Creek maintained sizeable areas with high levels of homogeneity of substrate in comparison to smaller stream sites Aarons and Little Wapiti creeks (Figures 1, 2). In this case, heterogeneity is to some degree a matter of scale, as habitat patches likely increase in size as the stream size increases. However, Elk River substrate deposition pattern, while relatively heterogeneous in some areas, does have a large area of sand interspersed with boulders and cobble. That model r^2 means of large and small stream groups were not significantly different even with marked difference of substrate depositional pattern is a strong indication of the effectiveness of GWR regression for modeling stream substrate. However, because decreases in accuracy of stream habitat models using GWR are caused by high levels of spatial heterogeneity attention should be directed towards the use of this GWR modeling method in highly heterogeneous streams, independent of stream size.

There are several future considerations brought into focus as a result of this study. Because of the correlation between depth and substrate and the ability of GWR to model it, future exploration should examine the potential of GWR for predictive modeling of substrate based on depth structure of the stream. Such an exploration would potentially remove the need for in depth substrate data collection once a baseline for modeling was created for the stream in question. It will be interesting to examine other frequently assessed variables and observe their response to similar methodology. This study also provides evidence that OLS residuals may guide map creation by demonstrating where models over or underestimate values.

In sum, consistent, accurate, and comparable spatial modeling of substrate on both large and small size wadeable streams are possible when using geographically weighted regression. Large and small stream substrate models responded equally as well to GWR while providing practical, easy to interpret maps of the data and analysis results. Further, the outcome eliminates the need for multiple types of statistics to be used to model streams of different sizes. It additionally provides a much needed method for comparison of large and small streams in the form of an r^2 value (by the same procedure) which gives proportional accuracy of models. Our methods have merit for fisheries managers because they provide comparable, clear results which may be used to visualize substrate, a key habitat variable, useful for management of fish populations in large and small wadeable streams.

Figure 1. Little Wapiti and Grayling Creek locations with 30 random sample points applied within each site. Actual substrate and depth information is also included for reference. Geographically weighted regressions were performed in a 2.5 meter radius surrounding each random sampling point.

Figure 2. Aarons Creek and Elk River locations with 30 random sample points applied within each site. Actual substrate and depth information is also included for reference. Geographically weighted regressions were performed in a 2.5 meter radius surrounding each random sampling point. Elk River displays greater areas of substrate homogeneity than other sites.

Figure 3. Residuals of ordinary least squares regression (top of figure) and geographically weighted regression residuals (bottom of figure) of Little Wapiti and Grayling creeks. Visual inspection of standard deviations of geographically weighted regressions contain a much larger amount of results within 0.5 standard deviations, the result of non-static variance applied at individual locations.

Figure 4. Residuals of ordinary least squares regression (top of figure) and geographically weighted regression residuals (bottom of figure) of Aarons Creek and Elk River. Elk River has large areas of more homogeneous substrate, which GWR was better able to model using local variance calculations.

Little Wapiti Creek Site

Grayling Creek Site

Aaron's Creek Site

Elk River Site

Little Wapiti Creek Site

Grayling Creek Site

Aaron's Creek Site

Elk River Site

**Standard Deviation
Of Residuals**

- < -2.5 Std. Dev.
- -2.5 - -1.5 Std. Dev.
- -1.5 - -0.5 Std. Dev.
- -0.5 - 0.5 Std. Dev.
- 0.5 - 1.5 Std. Dev.
- 1.5 - 2.5 Std. Dev.
- > 2.5 Std. Dev.

Literature Cited

- Ali, K., M. D. Partridge, and M. R. Olfert. 2007. Can geographically weighted regressions improve regional analysis and policy making? *International Regional Science Review* 30(3):300.
- Allan, J. D., and M. M. Castillo. 2007. *Stream ecology: structure and function of running waters*. Springer Verlag.
- Austin, M. 2007. Species distribution models and ecological theory: a critical assessment and some possible new approaches. *Ecological Modelling* 200(1-2):1-19.
- Borcard, D., P. Legendre, and P. Drapeau. 1992. Partialling out the spatial component of ecological variation. *Ecology* 73(3):1045-1055.
- Bradford, M. J., and J. S. Heinonen. 2008. Low flows, instream flow needs and fish ecology in small streams. *Canadian Water Resources Journal* 33(2):165-180.
- Brunsdon, C. 1995. Estimating probability surfaces for geographical point data: an adaptive kernel algorithm. *Computers & Geosciences* 21(7):877-894.
- Brunsdon, C., S. Fotheringham, and M. Charlton. 2000. Geographically weighted regression as a statistical model. Unpublished working paper, Department of Geography. University of Newcastle, Newcastle, UK.
- Charlton, M., S. Fotheringham, and C. Brunsdon. 2005. Geographically weighted regression. *NCRM Methods Review Papers* (NCRM/006).
- Chisholm, I. M., W. A. Hubert, and T. A. Wesche. 1987. Winter stream conditions and use of habitat by brook trout in high-elevation Wyoming streams. *Transactions of the American Fisheries Society* 116(2):176-184.

- Fotheringham, A. S., M. Charlton, and C. Brunsdon. 1996. The geography of parameter space: an investigation of spatial non-stationarity. *International Journal of Geographical Information Science* 10(5):605-627.
- Gao, X., Y. Asami, and C. J. F. Chung. 2006. An empirical evaluation of spatial regression models. *Computers and Geosciences* 32(8):1040-1051.
- Gergel, S., Y. Stange, N. Coops, K. Johansen, and K. Kirby. 2007. What is the value of a good map? An example using high spatial resolution imagery to aid riparian restoration. *Ecosystems* 10(5):688-702.
- Gido, K. B., J. A. Falke, R. M. Oakes, and K. J. Hase. 2006. Fish-habitat relations across spatial scales in prairie streams. Pages 265–285 *in* R. M. Hughes, L. Wang, and P. W. Seelbach, editors. *American Fisheries Society Symposium* 48. American Fisheries Society, Bethesda, Maryland.
- González-Megías, A., J. M. Gómez, and F. Sánchez-Piñero. 2005. Consequences of spatial autocorrelation for the analysis of metapopulation dynamics. *Ecology* 86(12):3264-3271.
- Gorman, O. T., and J. R. Karr. 1978. Habitat structure and stream fish communities. *Ecology* 59(3):507-515.
- Gotway, C. A., and L. J. Young. 2002. Combining Incompatible Spatial Data. *Journal of the American Statistical Association* 97(458):632-648.
- Hankin, D. G., and G. H. Reeves. 1988. Estimating total fish abundance and total habitat area in small streams based on visual estimation methods. *Canadian Journal of Fisheries and Aquatic Sciences* 45(5):834-844.
- Hansen, W. F. 2001. Identifying stream types and management implications. *Forest Ecology and Management* 143(1-3):39-46.

- Harvey, B., and A. Stewart. 1991. Fish size and habitat depth relationships in headwater streams. *Oecologia* 87(3):336-342.
- Ian, M. 1999. The importance of physical habitat assessment for evaluating river health. *Fresh Water Biology* 41(2):373-391.
- Knighton, D. 1998. *Fluvial forms and processes: A new perspective*. Oxford University Press Inc., New York.
- Koenker, R. 1981. A note on studentizing a test for heteroscedasticity. *Journal of Econometrics* 17(1):107-112.
- Kupfer, J., and C. Farris. 2007. Incorporating spatial non-stationarity of regression coefficients into predictive vegetation models. *Landscape Ecology* 22(6):837-852.
- Legendre, P. 1993. Spatial autocorrelation: trouble or new paradigm? *Ecology* 74(6):1659-1673.
- LeSage, J. P. 2001. A family of geographically weighted regression models. Pages 241-263 *in* L. Anselin, R. J. Florax, and S. J. Rey, editors. *Advances in Spatial Econometrics*. Springer, Berlin, Germany.
- Levin, S. A. 1992. The problem of pattern and scale in ecology: The Robert H. MacArthur Award Lecture. *Ecology* 73(6):1943-1967.
- Minns, C. K. 2001. Science for freshwater fish habitat management in Canada: current status and future prospects. *Aquatic Ecosystem Health & Management* 4(4):423-436.
- Moran, P. A. P. 1953. The statistical analysis of the Canadian lynx cycle. I. Structure and prediction. *Australian Journal of Zoology* 1(1):163-173.
- Rastetter, E. B., and coauthors. 1992. Aggregating fine-scale ecological knowledge to model coarser-scale attributes of ecosystems. *Ecological Applications* 2(1):55-70.

- Rosenfeld, J. 2003. Assessing the habitat requirements of stream fishes: an overview and evaluation of different approaches. *Transactions of the American Fisheries Society* 132(5):953-968.
- Scott, J., P. J. Heglund, and M. L. Morrison. 2002. *Predicting Species Occurrences. Issues of Accuracy and Scale*. Island Press, Washington, DC.
- Tilman, D., and coauthors. 2001. Forecasting agriculturally driven global environmental change. *Science* 292(5515):281.
- Townsend, C. R., S. Dolédec, R. Norris, K. Peacock, and C. Arbuckle. 2003. The influence of scale and geography on relationships between stream community composition and landscape variables: description and prediction. *Freshwater Biology* 48(5):768-785.
- Urban, D. L. 2005. Modeling ecological processes across scales. *Ecology* 86(8):1996-2006.
- Vannote, R. L., G. W. Minshall, K. W. Cummins, J. R. Sedell, and C. E. Cushing. 1980. The river continuum concept. *Canadian Journal of Fisheries and Aquatic Sciences* 37(1):130-137.
- Vitousek, P. M., H. A. Mooney, J. Lubchenco, and J. M. Melillo. 1997. Human domination of Earth's ecosystems. *Science* 277(5325):494.
- Wang, L., J. Lyons, P. Kanehl, and R. Bannerman. 2001. Impacts of urbanization on stream habitat and fish across multiple spatial scales. *Environmental Management* 28(2):255-266.
- Wheatley, M., and C. Johnson. 2009. Factors limiting our understanding of ecological scale. *Ecol Complexity* 6:150 - 159.

- Windle, M. J. S., G. A. Rose, R. Devillers, and M. J. Fortin. 2009. Exploring spatial non-stationarity of fisheries survey data using geographically weighted regression (GWR): an example from the Northwest Atlantic. *ICES Journal of Marine Science* 67(1):145-154.
- Winemiller, K. O., A. S. Flecker, and D. J. Hoeinghaus. 2010. Patch dynamics and environmental heterogeneity in lotic ecosystems. *Journal of the North American Benthological Society* 29(1):84-99.
- Zhang, L., and J. H. Gove. 2005. Spatial assessment of model errors from four regression techniques. *Forest Science* 51(4):334-346.

Chapter 7- Implications of the Research

Habitat assessment and modeling of streams and rivers is a collaborative process involving statistics, field methodology, knowledge of streams and stream ecology. Within the past decade and a half, assessment and modeling of streams has rapidly evolved to incorporate spatial data and complex mapping into its science. Since fall of 2004, we have investigated techniques to develop new avenues for modeling streams and streamline the process of their assessment through exploration of spatial habitat data on four wadeable streams, two each in West Virginia (Elk River and Aarons Creek) and Montana (Little Wapiti and Grayling creeks).

Initial exploration of stream habitat spatial data from Aarons Creek, West Virginia demonstrated potential of interpolation to provide practical maps using less data than standard grid sampling methodology (Sheehan and Welsh 2009). Following the study on Aarons Creek, West Virginia, three streams were added to the dataset; Elk River, West Virginia, Grayling Creek, Montana, and Little Wapiti Creek, Montana. Incorporating spatial qualities of substrate, depth, and flow velocity into a dataset by using geographically weighted regression provided a more accurate model of those variables in comparison to traditional linear regression. Results indicated failure to incorporate spatial qualities of data into a stream habitat study may provide ambiguous results and conceal even strong relationships between variables. Therefore we recommend using spatial statistics such as geographically weighted regression to model, map, and address stream habitat assessment because of the inherently spatial nature of stream variables.

Multiple datasets are often collected for stream ecological studies and habitat assessments because insights may be gained by looking at systems across varying spatial scales. In some

instances, this may no longer be a required action, as indicated by retention of interpolation accuracy of substrate at scales far removed from the scale at which the data was initially collected. In our study on Grayling and Little Wapiti creeks, sufficient accuracy of interpolations remained at scales 11 times larger than the initial data collection scale. This insight provides avenue for elimination of excess data collection for stream assessment and monitoring purposes. In stream assessment studies involving spatial data, such as mapping and assessment, a single dataset would be enough to provide guidance for future studies and management of the stream.

Heterogeneity is a frequent spatial attribute of stream habitat variables which occurs across spatial scales. The following insights were gained when data from Wapiti and Grayling creeks containing various levels of heterogeneity were analyzed using geographically weighted regression. First, heterogeneity of stream habitat has a negative effect on accuracy of this type of regression. Second, the more spatially heterogeneous the variable being modeled, the more data points needed to model it accurately. Third, geographically weighted regression may be used in conjunction with measures of spatial heterogeneity in streams much like Moran's I, to determine the amount of spatial heterogeneity a specific area of stream contains. These insights have the following implications: When modeling a stream it is important to quantify the amount of heterogeneity within the study area. Once the amount of heterogeneity is measured, potential efficiency of geographically weighted regressions may be known. This is important because too great a level of heterogeneity may render stream habitat models using regression analysis unusable. In addition, knowledge of how spatial heterogeneity may effect regressions will also guide the amount of data needed for accurate mapping and assessment using spatial regression.

Modeling, assessment, and comparison of substrate on small and large streams pose a variety of issues for fisheries scientists. Issues confronting accurate assessments across stream

size include incompatibility of spatial data collected at different scales and necessity of using more than a single type of statistical model, and failure to account for substrate complexity as stream size varies. These issues may result in inability to compare large and small streams, creation of inaccurate maps, ambiguous stream management results, and inefficient use of resources. To address these issues, chapter four examined the ability of geographically weighted regression to consistently model stream substrate on both large and small wadeable streams at an equivalent accuracy level. Our large and small stream substrate models responded equally as well to GWR while providing practical, easy to interpret maps of the data and analysis results. It additionally provided a much needed method for consistent comparison of substrate between large and small streams.

The type and amount of data often dictate the quality of the final product in stream habitat assessment and management as well as the type of analysis possible. While chapter two explored the repercussion of different sampling patterns on its analysis, chapters 3-6 examined analysis of previously collected spatial data. Random sampling methodologies for collection of stream habitat data operate under the correct assumption that the more data collected, the better the result. However, our results demonstrate a non-random sampling methodology is more efficient for mapping and assessment purposes than both random and grid sampling techniques. Increase in efficiency is accomplished by selecting data along habitat transition zones which have more spatial information and therefore more value associated with it than its adjacent neighbors. By actively selecting for more valuable data, interpolations and regressions require less information to achieve an acceptable level of accuracy for stream habitat study and management purposes.

The work found within the chapters of this dissertation, when combined, offers avenues to collect, model, and efficiently use stream habitat spatial data beyond the scope available through currently used methods alone. The chapters were designed to complement one another, thus providing a suite of new information for stream management and modeling. For example, Chapter six closely compliments the results from chapter three by providing an avenue for addressing spatial scale issues encountered when modeling and assessment of streams is undertaken not addressed by looking at data collection scale. When taken together, results of the two studies allow for less data collection, while still maintaining accuracy of stream studies and maps.

Management of streams is a delicate balance of knowledge and resources used to gain that knowledge. Assessment of stream habitat is one of the main tools used by managers to guide research and provide scientific insight. Our results have highlighted a few of the inadequacies in spatial modeling, and provided potential solutions. However, spatial modeling of streams is quickly evolving and as one solution is created, another problem arises to take its place. While our dataset was expansive, the study encompasses four streams, and limited aspects of those streams. Future work should focus on spatial modeling of additional streams in different ecosystems; it should encompass other emerging spatial statistics and explore existing ones more deeply. Management and modeling implications from our work could be strengthened and expanded with additional data and study of streams over time. In time, it is possible to envision incorporation of stream habitat data to develop a species specific model occurrence and population size estimate based on mapped and available habitat. When viewed from a broad standpoint, each facet of the dissertation is a method for reducing the data needed for stream monitoring and assessment while providing accurate, practical maps of study sites. In doing so,

the studies found within this dissertation allow for more efficient use of resources required for this endeavor, a desired result for anyone dealing with a finite budget of time and money. It is perhaps the greatest sum contribution of the work.

Appendix

Additional West Virginia Study Sites Detail

Insets Show Location and Aerial Location of Study Sites in West Virginia

The Elk River site measured 27 meters long by 22 meters wide.

The Aarons Creek site measured 24 meters long by 9 meters wide.

Location of Sampling
at Each Site

Additional Montana Study Sites Detail

Insets Show Location and Aerial Location of Study Sites in Montana

KENNETH R. SHEEHAN

13163 West Cross Place, Littleton CO 80127
(484)919-0624 • ken.r.sheehan@gmail.com

EDUCATION

- PhD - West Virginia University. 2011. Wildlife and Fisheries Resources. Dissertation topic: Examination of spatial modeling of stream habitat data.
- MS - West Virginia University. 2006. Wildlife and Fisheries Resources. Thesis; “An interpolation method for stream habitat assessments with reference to the Crystal darter (now Diamond darter).”
- BA - St. Olaf College. 1996. Triple major- English, Biology, and independent thesis study: “Man and nature: Examination of interactions from aspects of science, art, and philosophy.”

PROFESSIONAL EXPERTISE & EXPERIENCE

Spatial and GIS Based Analysis and Fisheries Expertise

- Geographic Information Systems data management, analysis, and map making expertise (ArcGIS 10).
- Spatial modeling of environmental data with a focus on aquatic and forest systems.
- Environmental and forest ecosystems remote sensing data acquisition and analysis.
- Freshwater aquatic systems, fish ecology, fish habitat and life history study and analysis.
- Data collection and sampling methodologies for game and non-game aquatic species.
- Spatial analysis and modeling of the environment across varying spatial scales.
- Suitability modeling.
- Database management.
- Image analysis.
- Spatial scale and model accuracy issues.
- Spatial autocorrelation and non-stationarity spatial model impacts and solutions.
- Minimal data spatial modeling techniques for assessment and analysis.
- Creative application of remote sensing and in field data acquisition to environmental modeling and assessment.
- Remote sensing and advanced spatial analysis instructor including LIDAR and sonar (bathymetry map creation and data collection).
- Image analysis.
- Addressing issues of scale and data incompatibility in spatial models.
- Integrating multidisciplinary environmental data in predictive models.

Statistics Topic Knowledge

- Spatial data analysis methodology- Ecological and non-ecological.
- Geographically weighted regression and spatial regression techniques.
- Interpolation techniques (such as trend, spline, natural neighbor, ordinary and universal kriging, co-kriging).

- “R” and “SAS” based statistical analysis.
- Maximum entropy (MAXENT) modeling.
- Spatial econometrics.
- Non-parametric statistics.
- Impacts of spatial autocorrelation on environmental study and analysis.
- AIC.
- Maximum entropy modeling.
- Value of information theory.
- Population dynamic analysis and study design.
- Quantitative ecological spatial analysis.

Research, Education & Professional Interests

- Building a foundation of education, research, and conservation which extends beyond my immediate influence.
- Assessment, monitoring, and prediction of change in spatial systems.
- Ecosystem modeling of terrestrial and aquatic ecosystems.
- Alpine and high elevation monitoring and change.
- Statistical analysis of spatial and non-spatial environmental data.
- Communicating cutting edge science to a broader public audience (beyond the classroom).
- Exploring values of land ethic and conservation and their impacts in science and science education.
- Applying value of information theory to data collection, assessment and analysis.

PROFESSIONAL EXPERIENCE

GIS & Research & Projects

- 2011- Cheat Lake, WV bathymetry map creation and habitat study.
- 2008-2011- Geographically weighted regression of stream habitat data modeling project. Spatial analysis of habitat on streams in the Greater Yellowstone Ecosystem.
- 2008-2011- Spatial heterogeneity study on frequently assessed stream habitat variables.
- 2008-2011- Spatial autocorrelation and non-stationarity study on spatial modeling of stream habitat variables.
- 2007-2010- Research assistant under Dr. Stuart Welsh. Long-term monitoring and assessment of eight baseline rivers in West Virginia.
- 2007- Brook trout study. Linkage between surficial geology and occurrence
- 2005-2007- Fish Habitat Assessment; minimal data interpolation methodology development
- 2004-2007- Development and testing of portable benthic fyke nets, capture of Diamond darter, habitat spatial study. Fish temporal movement study.
- 2004-2007, 2011- Diamond darter (rare fish) spatial habitat analysis, modeling, capture, and study.
- 2004-2007- West Virginia Cooperative Fish and Wildlife Research Unit (WVCFRU) under Assistant Unit Leader Dr. Stuart Welsh. Project: Diamond Darter collection and research (rare fish study).

Teaching

Full Courses

2010, 2011- Forestry 326, Remote sensing of the environment. Introducing students to GIS, LiDAR, and other remote sensing data acquisition and analysis methods for environmental and forestry purposes. Two, one-hour lectures and one, two-hour instructional lab each week.

Special Topic Seminars & Instruction

2010- Advanced spatial analysis (RESM 575). “Geographically weighted regression and other spatial techniques and their advantages for analysis of fisheries and environmental data.” Developed and taught three hour lecture and accompanying GIS statistical exercise. Additional focus on the importance of spatial autocorrelation in environmental study and analysis.

2010, 2009- Resource Management/Fisheries 440 (graduate level class); GIS for environmental applications. Two sections of 25 students. Weekly duties included lecture, demonstration and hands on teaching of spatial analysis methods and GIS applications for environmental research. Focus on watershed, fisheries, and wildlife uses of GIS.

2010, 2009, 2008- Biology 115 TA. Two class sections with lecture; Weekly duties included lecture, demonstration of biological principles, grading, and mentoring students on a weekly basis.

2010, 2009, 2008- Forestry 470B. “The art of fly fishing and the relationship between predator and prey.” An interactive lecture and discussion on the art of fly fishing and the effective choice and presentation of various flies using fish ecology concepts.

2009- Biology 104- TA with lecture; Weekly duties included teaching four classes, 20-30 minute lectures each including test and quiz preparation, demonstration of biological principles, grading (general requirement) 100 students on a weekly basis.

2007- Excel data analysis seminar. Advanced and beginning course. Developed, organized, and taught (sole instructor) multiday Excel seminar for West Virginia Department of Natural Resources.

2006- Ichthyology. “History of Ichthyology.” Guest lecture on chronology and development of personalities and innovation in ichthyology focusing on North America since the time of colonization.

TECHNICAL REPORTS & PUBLICATIONS

Sheehan, K.R., Welsh, S.A., and Strager, M.P. (Accepted October 2011) Advantages of geographically weighted regression for modeling benthic substrate in two greater Yellowstone ecosystem streams. *The Journal of Environmental Modeling and Assessment* (Will be online and in print in 2012).

Sheehan, K.R., Welsh, S.A., and Strager, M.P. (Accepted November 2011) Accuracy of stream habitat interpolations across spatial scales. *Journal of Geographic Information System* (Will

be online and in print in 2012).

- Welsh, S.A. and Sheehan, K.R. 2010. Assessment of long-term monitoring stations: Final report. West Virginia DNR. Pages 1-81.
- Sheehan, K.R., and S.A. Welsh. 2009. An interpolation method for stream habitat assessments. North American Journal of Fisheries Management 29(1):1-9.
- Welsh, S.A., R.M. Wood, and K. R. Sheehan. 2009. Threatened fishes of the world: *Crystallaria cincotta*, Welsh and Wood 2008 (Percidae). Environmental Biology of Fishes 84:191-192.
- Sheehan, K.R. 2006. An Interpolation Method for Stream Habitat Assessments with Reference to the Crystal Darter. (Master's Thesis) West Virginia University.

Currently in review

- Sheehan, K.R., Welsh, S.A. An Effective Data Selection Pattern for Modeling Benthic Substrate in Streams; Using Ordinary Least Squares Residuals to Guide Data Collection.
- Sheehan, K.R., Welsh, S.A., and Strager, M.P. Geographically Weighted Regression Models and Role of Morans I In Evaluating Heterogeneous Benthic Substrate In Streams.
- Sheehan, K.R., Welsh, S.A. Use of Geographically Weighted Regression to Model Benthic Substrate; A Comparison Between Large and Small Wadeable Streams.

PROFESSIONAL PRESENTATIONS

- 2010- American Fisheries Society National Meeting, Pittsburgh, PA, "Advantages of geographically weighted regression for modeling stream habitat."
- 2010- West Virginia Professional GIS Conference, Martinsburg, WV. "Advantages of geographically weighted regression for modeling substrate in streams." (Second place, best student paper).
- 2009- American Fisheries Society National Meeting, Nashville, Tennessee, "Stream habitat interpolation using minimal data."
- 2009- West Virginia University Annual Graduate Colloquium, Davis College of Consumer Sciences – "Interpolation methods using minimal data in GIS."
- 2008- USGS Headquarters Seminar, Reston, VA. "An interpolation method for stream habitat assessments: Minimal data predictive modeling in GIS."
- 2008- West Virginia American Fisheries Society Student Chapter Presentation - "The importance of internships & four qualities of successful scientists."
- 2007- West Virginia University seminar presentation, "Accuracy trends in predictive river modeling in GIS; The art of balancing data collection effort and accuracy gains."
- 2007- West Virginia Fish and Wildlife Research Unit annual meeting – "An interpolation method for stream habitat modeling and assessments."
- 2006- West Virginia University seminar presentation - "Elk River Crystal darter (*Crystallaria asprella* sp.) Ongoing research and capture methods using a variety of sampling gear."
- 2006- Southern Division AFS Annual Meeting, San Antonio, TX – "Habitat Modeling and Interpolation of Depth and Substrate on Aaron's Creek, West Virginia."
- 2005- Fluvial Geomorphology Seminar, West Virginia University – "Geological influences on Substrate values at Clendenin Shoal on Elk River, Kanawha County, West Virginia."
- 2005- Quantitative Ecology presentation, West Virginia University – "Validation of three types of interpolation for substrate class on a reach in Aaron's Creek, WV."
- 2005- Southern Division AFS Student Colloquium, Greenville, NC – "Habitat Modeling and

Interpolation Using GIS on Aaron's Creek, West Virginia.”
 2005- West Virginia University – “The Crystal Darter (*Crystallaria asprella*) life history and opportunities for research on Elk River, West Virginia.”

APPLICABLE COURSEWORK

Spatial Analysis for Resource Management	GIS: Technical Issues
Statistical Methods I	Advance Spatial Analysis Methods
Statistical Methods II	VBA Programming in ArcObjects
Non-parametric Statistics	Research Design
Quantitative Ecology	Critical Review: Fisheries
Advanced Fluvial Geomorphology	Grants and Grantsmanship
Wildlife Management	Advanced Wildlife Population Ecology
Fish Ecology	Teaching Practicum
Law- Environmental Policy (EPA, ESA, etc.)	Advanced Ichthyology
Forestry	Animal Physiology
Biology	Microbiology
Native Cultures in the American Southwest	Wildlife Seminar

SERVICE & PROFESSIONAL ASSOCIATIONS

2004-2011 - American Fisheries Society member.
 2010/2011 - West Virginia Association of Geospatial Professionals.
 2008-2011 - Extracurricular mentoring students in Geospatial technology studies and ArcGIS.
 2010, 2009, 2008, 2007, 2006, 2005 - American Fisheries Society annual fundraiser - Involving local businesses, collecting and developing prize pool, event organization, photography donations.
 2009 - US Forest Service, National Meeting of Experimental Forests - Invited PhD session aide, session recording and transcription, audio/visual and presentation organization and function.
 2009 - Yellowstone Park Foundation member.
 2007, 2008 - Editor, contributing author, West Virginia American Fisheries Society Chapter Newsletter.
 2007 - USGS Headquarters, Reston, VA. Open House - Educating the public about the USGS involving schools and NGO organizations in the greater Washington, DC area.
 2005/2006 - American Fisheries Society Student Chapter President.
 2005 - American Fisheries Society Annual Southern Division Meeting - Organization and planning of annual Student/Mentor lunch for 200+.
 2004, 2005 - AFS Annual Southern Division Meeting - Organization of student paper judges and awards for all student papers.
 1996 - Trail maintenance, Chippewa National Forest, Minnesota.
 1994 - Student Conservation Association (SCA). Sol Duc Ranger District, Olympic National Forest: *Spotted Owl* Nesting area research and location, *Marbled Murrelet* location and identification.
 1990 - Student Conservation Association (SCA). Grand Teton National Park, WY- Trail restoration.

1989 - John Pennekamp National Marine Sanctuary. Key Largo, FL- Volunteer Internship: Specimen collection, visitor education.

AWARDS & ESEI TEACHING REVIEWS SUMMARY

- 2011-Doctoral finishing grant- Awarded to PhD candidates based on academic merit and performance.
- 2011- Current ESEI Average (out of 5.0): 4.74 or 94.8 %. ESEI is a national standard which compares teaching evaluations and scores in comparison to other teaching professionals.
- 2010- Hoyt Teaching Fellowship.
- 2010- West Virginia Professional GIS conference, 2nd place, best student paper.
- 1996- St. Olaf College. One of five nominees for best senior project (Thesis). Three part study of nature and human interaction including Columbia River steelhead restoration project assessment, book of poetry exploring the relationship of man and nature from a creative perspective, and the book "Yellowstone Experience" (unpublished work) about solo winter camping alone in the Yellowstone backcountry.

OTHER PROFESSIONAL (APPLICABLE BUSINESS) EXPERIENCE

1998-2003 Advertising and Public Relations-

- Organized fund raisers and media events.
- In depth knowledge of non-profit organization functionality.
- Client relations for 200+ monthly advertisers.
- Prior responsibility for 50k monthly advertising revenue including international client list.
- Deadline oriented working environment.
- Grant writing.
- Worked closely with community leaders, media outlets, businesses in varying fields.
- Graphic design and promotional information creation.
- Creative writing and slogan development.
- Management of large events (movie premieres, restaurant openings, press conferences).

ACCREDITATIONS

- Red Cross First Aid and CPR certified (current)

REFERENCES

Dr. Stuart A. Welsh, Assistant Unit Leader - Fisheries (304) 293-5006
West Virginia Cooperative Fish and Wildlife Research Unit
West Virginia University
PO Box 6125, Morgantown, WV 26506
swelsh@wvu.edu

Dr. Patricia Mazik, Unit Leader (304) 293-4943
West Virginia Cooperative Fish and Wildlife Research Unit
West Virginia University
PO Box 6125, Morgantown, WV 26506
pmazik@wvu.edu

Dr. Mike Strager, Associate Professor (304)-293-6463
Resource Management
West Virginia University
PO Box 6108, Morgantown, WV, 26506
mstrager@wvu.edu

Additional references available upon request