An interpolation method for stream habitat assessments with reference to the crystal darter

Kenneth Richard Sheehan

West Virginia University

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

Recommended Citation
AN INTERPOLATION METHOD FOR STREAM HABITAT ASSESSMENTS
WITH REFERENCE TO THE CRYSTAL DARTER

Kenneth Richard Sheehan

Thesis submitted to the
Davis College of Agriculture, Forestry and Consumer Sciences
at West Virginia University
in partial fulfillment of the requirements
for the degree of
Master of Science
in
Wildlife and Fisheries Resources

Stuart A. Welsh Ph.D., Chair
J. Steven Kite Ph.D.
Michael P. Strager Ph.D.

Division of Forestry

Morgantown, WV
2006

Keywords: Interpolation, Natural Neighbor, Aquatic Habitat Modeling, Kriging,
Habitat Assessment

Copyright 2006 Kenneth R. Sheehan
AN INTERPOLATION METHOD FOR STREAM HABITAT ASSESSMENTS WITH REFERENCE TO THE CRYSTAL DARTER

ABSTRACT

Kenneth R. Sheehan

*Crystallaria asprella* spp. (Elk River crystal darter) is documented only in Elk River, West Virginia. Little life history detail is known about this lotic, benthic-dependent species. Previous studies have led to a very basic understanding of the use of depth, flow and substrate of the crystal darter (*Crystallaria asprella*), and less is known about the Elk River crystal darter. Due to substrate-specificity in crystal darters (which reportedly bury in sand), a study concerning benthic stream habitat modeling was undertaken. Substrate and depth modeling using >5% of an area sampled at a resolution of 0.093 m² square spatial data in a Geographic Information System (GIS) was theorized possible. Representation of actual substrate using 2268 0.093 m² data cells with UTM coordinates was created in ESRI ArcMap version 9.1. Each cell signified the dominant substrate type found in that area. We selected points representing 5% and 2.5% percent of the site area for both depth and substrate and natural neighbor interpolations were run on these points. The actual values of depth and substrate were compared with predicted values to determine accuracy of interpolated data. The 5% interpolations were more accurate for both depth and substrate than 2.5% sampling results. The 2.5% interpolations achieved accuracy up to 92% of actual values and interpolations based on 5% within 5% of actual when comparing area of substrate predicted. Depth predictions based on 2.5% attained accuracy from 49% to 92% when applied to threshold values while 5% percent interpolations illustrated accuracy levels ranged from 57% to 95% for the same thresholds. Our findings demonstrate the use of minimal amounts of fine-scale data of substrate and depth for interpolation of habitat in large areas of a stream channel. This approach allows time and cost saving options for sufficiently accurate microhabitat scale habitat assessments of large sections of rivers, and provides functional maps to aid habitat-based fisheries management.
DEDICATION

Thanks family, friends, and colleagues! Reality aside (no this is not sarcastic), it’s been a wonderful ride.
ACKNOWLEDGEMENTS

I would like to thank R. Utz, G. Staines, and your mom for numerous scientific conversations and idea sessions. Thanks as well to Pat Mazik for meeting with me initially and letting me join the coop (as well as funding). Kyle Hartman, thanks for the use of the computer in the GIS lab, without which, I wouldn’t have been able to finish this thesis.
TABLE OF CONTENTS

ABSTRACT.................................................................................................................................. ii
DEDICATION............................................................................................................................ iii
ACKNOWLEDGEMENTS......................................................................................................... iv
LIST OF TABLES .................................................................................................................. vi
LIST OF FIGURES ................................................................................................................. vii
CHAPTER 1: Literature Review ............................................................................................. 1
  Introduction Statement......................................................................................................... 1
  Background information for Elk River Crystal darter ........................................................ 2
  Stream habitat modeling .................................................................................................... 4
  Conclusion paragraph ....................................................................................................... 6
  Literature Cited ................................................................................................................ 7
CHAPTER 2: An interpolation method for stream habitat assessment with reference to the Elk River crystal darter ............................................................................................ 13
  Abstract............................................................................................................................. 13
  Introduction ....................................................................................................................... 14
  Methods .............................................................................................................................. 17
  Study area ....................................................................................................................... 17
  Interpolation of substrate and depth .................................................................................. 19
  Accuracy of interpolations ............................................................................................... 20
  Results .............................................................................................................................. 21
  Habitat mapping ............................................................................................................. 21
  Substrate interpolation ..................................................................................................... 21
  Depth interpolation ......................................................................................................... 21
  Discussion ......................................................................................................................... 22
  Acknowledgements ......................................................................................................... 25
  Literature Cited .............................................................................................................. 25
APPENDIX............................................................................................................................ 41
LIST OF TABLES

Table 1. USDA soil survey manual substrate size delineations used as guideline for substrate classification (Soil Survey Division Staff 1993) .......................................................................................................................... 31

Table 2. Substrate root mean square error values of complete substrate based on actual values .............................................................................................................................................................................32

Table 3. Depth interpolation root mean square errors based on complete depth (actual values), 5% interpolations and 2.5% interpolations .........................................................................................................................33
LIST OF FIGURES

Figure 1. Simple flow chart containing general overview of data collection and interpolation process ..................................................................................................................32

Figure 2. Substrate raster interpolation maps comparing actual (known) values to interpolations based on 5% (A) and 2.5% (B) of actual values ........................................................................................................36

Figure 3. Depth raster interpolation maps comparing actual (known) values to interpolations based on 5% (A) and 2.5% (B) of actual values .....................................................................................................37

Figure 4. Natural neighbor substrate interpolations based on actual (2268 points.), 5% of actual values (A; 113 points), and 2.5% of actual values (B; 56 points) .........................................................................................................38

Figure 5. Depth Percent match of interpolation A (based on 5% of actual values) and B (based on 2.5% of actual values) when compared to initial actual values..........................................................................................39

Figure 6. Depth interpolation matching (all) exact depth values within 5, 10 and 20 % thresholds (5% threshold = % depth values within 3cm of actual values).................................................................................40
CHAPTER 1: Literature Review

Introduction statement

Little is known about the Elk River crystal darter (*Crystallaria asprella* spp.), a rare species that is known from 12 specimens collected during 1980-2005 (Cincotta and Hoeft 1987, Osier 2005, Warren et al. 2000). The objectives of this research were to document movement and habitat use of the Elk River crystal darter, but the inability to capture specimens hindered my research efforts. Researchers have targeted Elk River crystal darters with multiple capture gears including snorkeling, straight seines, bag seines, trawls, boat electrofishers, backpack electrofishers, and parallel wires electrofishers. Of these gears, straight seines, bag seines, and boat electrofishers were successful, but with an extremely low catch per unit effort that indicates either a small population or an inefficiency of gear types. Consequently, I designed a small fyke net for use in medium sized rivers to try a different gear type. Fyke nets were fished in riffle, glide, and head of pool habitats during expected crepuscular and night movements of crystal darters from pool to riffle habitats. These movements were suggested for other populations of crystal darter (Grandmaison et al. 2003, Osier 2005, Hatch 1997). The fyke net sampling did not capture crystal darters, and results of its effectiveness on capturing other species will be reported separately. I was also interested in habitat use of the Elk River crystal darter, but only captured two individuals via 50 foot bag seine during my study. Consequently, I refocused my efforts toward developing a computer-based interpolation method of modeling stream habitat. The modeling approach was applied to data from a small stream, but should be applicable to larger rivers due to similar depositional processes across scales, such as the Elk River, and allows
researchers to accurately estimate stream habitat from a small sample of data points. Further, the habitat modeling method allows researchers to predict the amount of available habitat for a given species given previous knowledge of habitat use requirements.

**Background information for Elk River Crystal darter**

The Elk River crystal darter is a rare fish currently limited to several pool and riffle locations in a 30 km stretch in Elk River, West Virginia between Sutton Reservoir and Charleston, WV (Cincotta and Hoeft 1987, Osier 2005). Currently *C. asprella* populations persist in Ohio, Missouri and Mississippi drainages though population numbers have declined. *Crystallaria asprella* is considered extirpated from other Ohio river drainage states Tennessee, Illinois, Indiana, Kentucky and Ohio (Cincotta and Hoeft 1987). Abundance is unknown due to lack of consistent capture. Sampling gear vulnerability of *C. asprella* and *C. asprella* spp. is minimal (Osier 2005, Grandmaison et al. 2003). A total of twelve Elk River crystal darters have been collected since 1980 in the Elk River drainage and the Elk River crystal darter is recognized as a vulnerable species (Warren et al. 2000). The four most recent individuals were collected two each in summer 2003 and summer 2005.

Crystal darters were originally described as *Pleurolepsis asprellus* (Jordan 1878). There is general agreement crystal darters fit into the monophyletic group of all darters and there is no current documentation indicating *C. asprella* ssp. falls outside this group. This darter group is further delineated into quad-generic classification between *Ammocrypta, Crystallaria, Etheostoma* and *Percina*. In 1897, Jordan reclassified the fish as *Ammocrypta asprella* which it remained for 94 years (Page 1981). Page (1981) used
the presence of 12-15 dorsal spines and 13-14 dorsal rays as evidence for recognizing the
genus *Crystallaria*.

The Elk River crystal darter population is an evolutionary significant unit (Wood
and Raley 2000, Morrison et al. 2006) originating from *C. asprella* stock whose historic
range spanned from West Virginia, west to Missouri, and from Minnesota south to the
Genetic analysis has shown Elk River *C. asprella* differs 10-12% in mitochondrial DNA
sampling of the cytochrome *b* gene in comparison to other crystal darter populations
(Wood and Raley 2000, Morrison et al. 2006). Recent close examination of *C. asprella*
spp. genetic variation from other crystal darter populations has yielded similar results and
recommendations (Morrison et al. 2006). This size of the genetic divergence has prompted
suggested listing of the Elk River population as either a subspecies or new species of
(2000) recognized the Elk River crystal darter as a subspecies. Future work in
phylogeography, an emerging field explaining genetic lineage with emphasis on
goingraphic location, may indicate the recent evolutionary ancestor of the Elk River crystal
darter (Waters et al. 2001).

Little is known about life history and habitat use of the Elk River crystal darter in
part due to difficulty in collection and sampling of large river benthic habitat and limited
field season (George et al. 1996, Osier 2005). The Elk River crystal darter is thought to
follow habitat use of other *C. asprella*. Typical collection depths of other *C. asprella*
range from 0.5 - 2.0 meters. The fish are rarely collected in velocities less than 32 cm/sec
(George et al. 1996, Osier 2005) which seems to indicate lower flow velocity preferences.
Elk River specimens have been caught over predominantly gravel substrate intermixed with boulders and sand patches (Cincotta and Hoeft 1987, Osier 2005). It is believed to mirror observed *Crystallaria* species behavior pattern of burying in sand (Miller and Robinson 1973). Due to capture specimen size typically in the range of 80-100 mm, life span is assumed to be similar to other *Crystallaria* with a life cycle of two years on average (Hatch 1997). Elk River crystal darters are closely associated with *Ammocrypta pellucida*, the eastern sand darter, though less translucent, more completely scaled and larger in size (Simons 1996).

There have been concerted efforts by E. Osier and S. Welsh with extensive bag seine episodes over a two year period spanning summers of 2003 and 2004. During this time period only two specimens were captured at the site in the town of Clendenin, WV (Osier 2005). Dan Cincotta a biologist in the WV Department of Natural Resources has conducted annual samplings of the Elk River with no additional specimen captures.

*Stream habitat modeling*

Stream river habitat is often viewed as a heterogeneous complexity of water currents, water depths, and rock sizes (Bain and Stevenson 1999, Rosgen 1996, Knighton 1998, Komar and Carling 1991, Leopold and Maddock 1953). Although many components contribute to stream habitat, researchers often view water current, water depth, and rock size as components perceived by fish as important (Kohler and Hubert 1999, Bain and Stevenson 1999, Arend 1999). In some smaller streams, habitat is relatively simple to measure, owing in part to accessibility, but larger streams pose difficulties in habitat assessment owing to the time it takes to survey larger areas and inaccessible areas of fast and deep waters. Stream habitat modeling, a computer-based
approach to estimating habitat, is a method that overcomes some of the difficulties with habitat assessments in larger rivers.

The predictability of stream habitat components allows one to effectively model habitats (Armstrong 2000). Stream habitats are predictable owing to patterns in particle size deposition relative to water current velocities (Smith and Ferguson 1995, Keller 1971, Knighton 1998). Substrate, velocity, and depths of stream habitats are structured at map scales, such as riffle/pool sequences (Leopold and Maddock 1953, Powell 1998) and microhabitat scales, such as rock size sorting (Komar and Carling 1991), and velocity-controlled depositional areas with abrupt substrate transitions (Smith and Ferguson 1995, Ferguson 2003). Results indicate the spatially-correlated structure of stream habitats promotes highly predictable and accurate interpolations of substrate and depth data.

With the advance of computer technologies, researchers have the opportunity to interpolate or estimate measurements within a geographic area based on a sample of data points of the area of interest (Fisher and Rahel 2004). Interpolation methods are used in many diverse fields of science, including forestry (He et al. 2000), ornithology (Dettmers and Bart 1999), medical science (Vine et al. 1997), weed science (Zille et al. 2002) and fisheries (Toepfer et al. 2000). Several types of interpolation approaches are available, each one with specific applications, including nearest-neighbor, ordinary kriging, inverse distance weighted (IDW), pointinterp (similar to IDW), spline (which minimizes surface curvature), and combinations thereof (co-kriging). Some methods of kriging may be manipulated to account for directional weighting, i.e. when downstream is oriented on the geographic plane, it is possible to “weight” interpolations accordingly.
Sand patches, often have a specific shape and edge boundary (Smith and Ferguson 1995, Ferguson 2003) and are a good example of directional weighting in nature. Search radius may also be altered, which allows interpolation to react to a variety of spatial scales.

We initially entertained using ordinary kriging, IDW, or nearest (natural) neighbor (NN) interpolations for this project. Inverse distance weight interpolation has often been compared to ordinary kriging (Meuller et al. 2004, Tabios and Salas 1985, Kravchenko and Bullock 1999, Zimmerman et al. 1999, Zille et al. 2002). Results of prior studies indicate both methods have promise, but which performs better is still a point of contention and seems to depend on validation method and type of data being interpolated (Meuller et al. 2004). The IDW method is not as good at interpolating data with spatial qualities as other methods (Meuller et al. 2004). The NN interpolation method is appropriate for stream habitat because predicted values of cells are heavily-influenced by values of adjacent cells (Sibson 1981). In streams, habitat structuring produces differentially-sized patches of specific habitat types (Boyero 2003), so that two samples of habitat within adjacent areas of a small area are likely to be similar or spatially correlated. Spatial autocorrelation is a concern for some spatial studies (Liebhold and Gurevitch 2002), but is an important component that contributes to predictability and accuracy in habitat interpolation methods (Robertson 1987).

Conclusion paragraph

This chapter reviewed literature relative to the Elk River crystal darter and stream habitat modeling. Little is known about the Elk River crystal darter, including a lack of knowledge about the use and availability of habitat in the Elk River. Quantifying habitat availability for long sections of large rivers, such as the Elk River, has previously
been cost-inhibited. In Chapter 2, we provide a method to interpolate stream habitat
and create accurate maps of river habitat using minimal real world spatial data.
Although the habitat interpolation methods were developed on a small stream, we
believe that the methods will apply to larger systems, and further research efforts will
focus on applying these methods toward habitat interpolation in Elk River, West
Virginia.

**Literature Cited**

Association of American Geographers 90:146-156.

Arend, K.K. 1999. Classification of streams and reaches. Pages 57-74 in M.B. Bain, and
N.J. Stevenson, editors. Aquatic habitat assessment: common methods. American
Fisheries Society, Bethesda, Maryland.

methods. American Fisheries Society, Bethesda, Maryland.

Boiero, L. 2003. The quantification of local substrate heterogeneity in streams and its


Dettmers, R. and J. Bart. 1999. A GIS modeling method applied to predicting forest

Press, Knoxville.


Chapter 2: An interpolation method for stream habitat assessments with reference to the Elk River crystal darter

Abstract

Data interpolation methods for habitat assessments reduce time and labor costs by using a small number of habitat samples to predict habitat of larger areas. The spatial correlation of stream habitat variables, such as substrate and depth, improves accuracy of interpolated data. Using a GIS (ArcMap version 9.1), nearest neighbor interpolation was used to predict substrate and depth (based on 2.5% and 5.0% of the total area) within a 210.7 m² section of a second order stream. The true values of depth and substrate were quantified for the entire study site and compared with predicted values to determine accuracy of interpolated data. The 5% interpolations were more accurate for both depth and substrate than 2.5% sampling results. The 2.5% interpolations achieved accuracy up to 92% of actual values and interpolations based on 5% within 5% of actual when comparing area of substrate predicted. Depth predictions based on 2.5% attained accuracy from 49% to 92% when applied to threshold values while 5% percent interpolations illustrated accuracy levels ranged from 57% to 95% for the same thresholds. Our findings demonstrate the use of minimal amounts of fine-scale data of substrate and depth for interpolation of habitat in large areas of a stream channel. This approach allows time and cost saving options for accurate microhabitat scale habitat assessments of large sections of rivers, and provides functional maps to aid habitat-based fisheries management.
Introduction

The measurement of habitat is central to the management and conservation of fishes (Bain and Stevenson 1999, Murphy and Willis 1999, Noss et al. 1997). Substrate type, water depth, and water current velocity are often considered as important stream habitat variables in both ecology (Gorman and Karr 1978) and fluvial geomorphology (Knighton 1998) and are often collected at fine (i.e., microhabitat) scales in ecological studies, such as one-meter square sampling areas or smaller (Greenberg 1991, Simonson 1993, Welsh and Perry 1998a, 1998b). Estimation of microhabitat variables over large geographic areas is typically cost prohibitive owing to time and labor-intensive methods. Because of time and labor requirements, managers sometimes opt for collection of large scale macrohabitat data for long sections of stream habitat, although the finer scale data of microhabitats may be more appropriate to address management needs and conservation concerns. Recent advances in GIS-based data interpolation methods allow prediction (based on a small number of sampling points) of fine scale spatial habitat data across large geographic areas. Data interpolation methods are increasingly used in natural resource studies (Dettmers and Bart 1999, He et al. 2000), and should also be a useful component of stream habitat assessments.

Substrate, velocity, and depths of stream habitats are structured at larger landscape scales, such as riffle/pool sequences (Leopold and Maddock 1953, Powell 1998) and smaller scales, such as rock size sorting (Komar and Carling 1991), and velocity-controlled depositional areas with abrupt substrate transitions (Smith and Ferguson 1995, Purkait 2002). The spatially-correlated structure of stream habitats promotes highly predictable and accurate interpolations of substrate and depth data.
Common deterministic and geostatistical interpolation methods are available for creating models and include inverse distance weighted (IDW), spline, ordinary kriging, pointinterp, natural neighbor (NN), and trend. Trend and spline minimize curvature of the interpolated surfaces, while IDW and pointinterp assign weights to cells as a function of distance. As such, IDW and pointinterp work well when data points are (more) dense and are able to delineate clearly a complex surface. Kriging utilizes statistical interpolation and is well suited to mapping directional influence on a surface with adequate data. There are subtypes of kriging including universal, ordinary and co-kriging which are based on similar statistics but have unique strengths. Universal kriging in particular is able to address the issue of directional influence in a data set.

Natural neighbor interpolation, which we use for this study, is interpolation based on weighted values of neighboring points. Mathematically, NN interpolation uses a Theissen polygon network to calculate those values. Resulting qualities of NN interpolation include the ability to handle large numbers of input points and handle clustered (non random and spatially auto-correlated) data well. Inverse distance weight and NN are closely related, using similar formulas for calculations. However, NN weights spatially closer known values more heavily. The NN interpolation method is appropriate for stream habitat modeling because predicted values of cells are heavily-influenced by values of adjacent cells. In streams, habitat structuring produces differentially-sized patches of specific habitat types, so that two samples of habitat within adjacent areas of a small area are likely to be similar or spatially correlated. Spatial autocorrelation is a concern for some spatial studies (Liebhold and Gurevitch...
2002), but is an important component that contributes to predictability and accuracy in habitat interpolation methods (Robertson 1987).

Fishery managers need accurate methods of habitat measurement and mapping with fewer constraints of time, budgets, and manpower (Arnason 1990) of which map creation will be a prominent aspect (Meaden 2004). Maps of stream habitat provide fishery managers with a template for understanding abiotic and biotic components, as well as a basis for development of management options. Managers benefit from interpolated habitat maps because less data (a savings of time and labor) can achieve similar results as traditional data collection methods. The cost savings associated with interpolated data are beneficial only if the predicted data are accurate. In general, a larger sample size of initial data points will increase the accuracy of interpolated data, but the relationship between sample size and accuracy is partly controlled by habitat heterogeneity (homogeneous habitats will require fewer initial data points to achieve accurate interpolations).

The study objective was to examine the use of predictive modeling of substrate and depth in large areas of a stream channel using natural neighbor interpolation based on minimal geographic data in a GIS. Our hypothesis revolves around the inherent belief that fisheries managers need faster yet accurate methods to evaluate river habitat. Interpolation in geographic information systems is a relatively new and powerful set of methods in which the extent of modeling methodology and success is just beginning to be explored for fisheries professionals. With GIS, for the first time in our field, fisheries data and science is linked with creation of dynamic and practical maps of habitat and species for management purposes. This ability allows us to approach new time and cost saving
options for accurate microhabitat scale habitat assessments of large sections of rivers, and provides functional maps to aid habitat-based fisheries management. Specific to this thesis, this method is being developed as an effective tool for purposes ranging from monitoring to management for the Elk River crystal darter (*Crystallaria asprella* ssp.), other benthic fishes, and mussels in Elk River, West Virginia.

**Methods**

**Study area**

Our study site was located on a reach of 8.4 mile long Aaron Creek of the Monongahela River system in Monongalia County, West Virginia. The 2004 West Virginia Department of Environmental Protection classified the stream category 4a - Impaired or threatened for one or more designated uses and TMDL has been completed (WVDEP 2004). The northwest corner of the sample site was (591540, 4386028) NAD 1983 UTM Zone 17N. There is sparse to moderate suburban and business development along parts of the stream, though water quality and fish diversity remain high even with the 4a classification. Riparian area along the stream is a mixture of field, lawn and hardwood forest between 5 m and 50 m in width. Canopy cover was calculated using a densiometer at the upstream and downstream borders of the site and averaged to 25% overall. This reach was selected because it contained a complete riffle and downstream pool which appeared to contain natural characteristics of flow and sinuosity and mirrored Elk River substrate and depth attributes. The site encompassed a surface area of 210.7 m².

An overview of the process used for this project includes data collection, geo-referencing of site location, import and preparation of data in MS Excel, use of data in
ArcMap for interpolation and modeling, and subsequent export and analysis of data (Figure 1). The spatial scale for the measurement of stream substrate was smaller than that used for stream depth because depth was more homogenous over a given area than substrate. For the measurement of substrate, the study site was divided into a grid of 2,268 0.093 m² cells. The common substrate size category was recorded within each cell (1 = silt, 2 = sand, 3 = gravel, 4 = cobble, 5 = boulder and 0 = land) based on the Soil Survey (1993) soil sample standards (Table 1, Appendix 1). After entering the site corner locations into ArcMap, they were geo-referenced for accuracy and adjusted to match measured site size. The sample quadrant size of 0.093 m² was smaller than our global positioning system accuracy (GPS), so adjusted site corner values were assigned (in an MS Excel spreadsheet) UTM coordinates for the center of each 0.093 m² cell (2,268 points) using the series fill function. Performing this function allowed even distribution of x,y coordinate points at the specified resolution and ensured that completed substrate and depth data sets were complete and accurate for import by a GIS (ArcMap 9.1). We have for this experiment located the site on the geographic plane with exact coordinates; while it is important to have the site adequately located on the geographic plane, the interpolations would be possible without an exact location so long as scale was kept the same within the GIS. In this situation, exact coordinates could be assigned later.

For measurement of depth, the site was divided into a grid of 1,188 0.177 m² cells. Depth was measured at the center of each cell using a continuous scale (cm), and cell values were imported to a GIS (ArcMap 9.1). The 2,268 substrate values and 1,188 depth values provided complete spatial coverages of the study reach. These complete
site coverages of substrate and depth values represented the real world site digitally (hereafter called the actual coverage, at aforementioned resolutions, of substrate or depth) and were used in accuracy assessments of interpolated values.

*Interpolation of substrate and depth*

The selection process of points for the 2.5% and 5% interpolations is key to successful interpolation map creation. Trial and error indicated a combination of edge boundary and feature center point selection was most effective in selecting points for our interpolations. Therefore, initial key points were selected at terminal edge boundaries, where (if on a grid placed across a stream reach) the feature ended on both the up and downstream and cross stream locations; the terminal end of a sand bar would be an example, where the bar does not continue downstream or across the location, i.e. the terminus. Once key terminus points were selected (including site corner boundaries), intermediate points were selected, in which a feature ended in two directions (up and downstream, or left and right cross stream) of which the end of a sand bar continuing downstream but having a border next to the thalweg would be an example. Once these points had been added to the layer being created for a specific % interpolation, center points of features were added showing continuation of a feature on all sides of a point (middle of a sandbar). Approximately 65% of selected points were located on intermediate and terminal boundaries, and the remaining 35% indicated center-points of features. A clustered pattern, which NN interpolation deals well with, is the end result of this procedure in which we sampled both 2.5% and 5% of cells from the actual coverage of substrate and depth.
In ArcMap, 113 locations representing 0.093 m² cells (i.e., 5% of dominant substrate composition at 0.093 m² resolution at the 210 m² site) was used for nearest neighbor interpolation of a new coverage of 2,268 substrate cells. Similarly, we interpolated another new coverage with a selection of 56 new cells from the same initial dataset (2.5% of the 2,268 total cells). For depth, we interpolated two new spatial coverages with samples of 30 (2.5%) and 59 (5%) cells of the 1188 cells from the site at a resolution of 0.177 m² coverage of depth data.

**Accuracy of interpolations**

The new data coverages of substrate and depth based on 2.5% and 5.0% interpolations were compared to values from the actual coverages. Substrate comparison was made between percent of area match for each substrate and total area for all substrates combined. Area in this case was defined by a 0.093 m² assigned to each cell the 2268 of which make up the total study site. Thus, match percentage was also performed, assessing when true and predicted values matched exactly. The accuracy of depth interpolation was assessed based on assigned threshold values. Site depths ranged from 0 to 60 cm, and we estimated 5%, 10%, and 20% threshold values which correspond to interpolated values within 3 cm, 6 cm, and 12 cm of the true value. Root mean square error (RMSe) was also calculated for interpolated depth and substrate values to compare to the digital representation of our site. To calculate RMSe we first calculated standard deviation for our sample and then used the formula \[ \frac{\sigma}{\sqrt{n}} \] where \( \sigma \) = standard deviation of the sample and \( n \) = sample size (number). Root mean square error was used because it indicates dispersion of data, and comparing dispersion levels of
interpolations is another indicator of which interpolation matches best with the digital representation of our study site’s depth and substrate values.

**Results**

*Habitat Mapping*

Raster maps of interpolated values of substrate and depth depicted close approximations to the study reach maps of true values (Figure 2 and 3). The raster map interpolated from the 5% sample more closely mirrors measured reality in structure and location of substrates and edge boundaries than the 2.5% interpolation (Figure 2). Similarly, the raster map interpolated from 5% of depth samples provided a more accurate representation of shape and overall depth structure of the study site than that from the 2.5% sample (Figure 3).

*Substrate Interpolation*

Based on the exact match between actual and interpolated substrate cells, the 5% and 2.5% samples correctly interpolated cell substrate categories on 61 and 46 percent of the total cells, respectively (Figure 4). Based on the percent area match between actual and interpolated substrate cells, the 5% and 2.5% samples correctly interpolated cell substrate categories on 79% and 54% of the total cells, respectively (Figure 4). When substrate categories were considered separately, less than a 5% difference in area occurred between actual cells and those interpolated from the 5% sample, with the exception of silt (Figure 5). Further, based on the 2.5% sample, a similar accuracy of less than 5% occurred for land, cobble, and sand categories, but differences were larger for silt (11%), gravel (15%), and boulder (19.5%; Figure 5). Root mean square error values for substrate supported our percent accuracy levels, showing that interpolated
values closer in dispersion (RMSe value) to actual values had higher accuracy levels (Table 2). Sand interpolation, which was most accurate, most closely matched actual RMSe values (Table 2).

Depth Interpolation

Depth data were interpolated with a sample of 2.5% and 5% of the total number of depth cells. Accuracy of depth interpolations was assessed with the percent of interpolated values within several ranges of actual values based on thresholds of 5, 10, and 20% of the maximum site depth of 60 cm (Figure 6). Based on interpolation from 5% of the depth measurements (59 data points), 57.3% of the interpolated depths were within a 5% threshold (3 cm) of the actual values, 82.7% percent within 10% (6 cm), and 95.1% of all depth values were within 20% (12 cm) of actual values (Figure 6). A lower accuracy occurred for interpolated values from 2.5% of depth measurements; 49% of values were within a 5% threshold (3 cm = 5% of maximum depth at site) of the actual depth values, 71% of interpolated values were within 6 cm of actual values, and 92.8% were within 12 cm of actual values (Figure 6). RMSe values reflected percent accuracy levels, and indicated that the 5% depth interpolation was closer in dispersion levels than the 2.5% interpolation (Table 3).

Discussion

Given the importance of habitat to the management and conservation of fishes (Noss et al. 1997, Orth and White 1999, Kohler and Hubert 1999, Thayer 1996), methods are needed for accurate, and cost-effective habitat sampling. Further, the technology to convert habitat data into accurate and usable habitat maps is an increasingly important benefit for aquatic species management (Kostylev et al 2001,
Manson and Todd 2000, Smith and Greenhawk 1998, Meaden and Do Chi 1996, Fisher and Rahel 2004). Our findings support a need for habitat interpolation methods in aquatic management and conservation programs, because a relatively small amount of data at fine spatial scales can be used in interpolation of fine scale data across larger habitat areas. This high accuracy of interpolated values and habitat predictability is possible owing to the spatially-correlated and structured features of stream habitats (Powell 1998, Rubec et al 1998, Jeffrey and Edds 1997, Bain et al. 1999, Ferguson 2003, Knighton 1998, Lisle 1979, Keller 1971, Leopold and Wolman 1957). This method provides fishery managers with accurate habitat assessments and habitat mapping with low constraints of time, budgets, and manpower.

Habitat assessments of long stream sections either focus on large area management objectives or accommodate cost constraints through avoidance of time and labor-intensive efforts of smaller scale data. Habitat assessments often focus on large areas, such as watersheds or stream reaches, but may also require fine-scale microhabitat data to address management needs and conservation concerns. Data interpolation methods can predict fine-scale spatial habitat data across large geographic areas, because habitat across large areas can be interpolated from data collected at a small number of sampling points (King et al. 1991, Rastetter et al. 1992). This approach allows interpretation of small scale (microhabitat) data over large geographic areas, and allows rasterized-data of microhabitats to be combined into larger scale habitat assessments. Conversely, data initially collected at large spatial scales are not easily or accurately scaled down to smaller areas (McPherson et al. 2006).
Spatial autocorrelation is a well-documented concern for geospatial studies (Henebry and Merchant 2002, Liebhold and Gurevitch 2002), but is an important component that contributes to predictability and accuracy in some ecological mapping and prediction methods (Klute et al. 2002, Rotenberry et al. 2002), including habitat interpolation methods (Robertson 1987). In stream habitats, proximate areas are more similar than distant areas, particularly with common stream characteristics of substrate, velocity, and depth (Leopold and Maddock 1953, Powell 1998). The nearest-neighbor interpolation method accurately predicted stream depth and substrate using small amounts of data.

This habitat interpolation method provides fishery managers with accurate habitat measurement and mapping with fewer constraints of time, budgets, and manpower (Arnason 1990). With this interpolation method, small amounts of data (a savings of time and labor) achieve similar results to large-scale time and labor intensive sampling efforts. The accuracy afforded by interpolation further promotes the cost-savings associated with interpolated data. Accuracy of this approach however is tied to habitat complexity and the spatial scale of data collection. Heterogeneous habitats will require a larger number of sampling points to achieve higher levels of accuracy; however, management decisions are often possible with low-resolution geospatial data or close approximations of the parameters of interest, such as habitat categories.

Although we examined accuracy of interpolated data based on initial 2.5% and 5% samples, accuracy is expected to improve with an increase in the initial sample size, and may also improve with a single analysis of combined data, such as interpolation with co-kriging of multiple habitat characters. Further research should examine the threshold scale
that relates sample size with accuracy (Host et al. 1995, Winkler and Fang 1997), and the related cost-benefit function that accompanies an increasing sample size. This relationship is also expected to differ with habitat complexity, and interpolation of stream habitat within pools would likely require a smaller initial sample size than interpolation of riffle habitats. At a larger scale of interest, interpolation of habitat for a stream section with a large pool/riffle ratio would likely require a smaller sample size relative to one with a complex high gradient profile (Keim and Skaugset 2002). Further analysis of interpolation with universal kriging and co-kriging, which both allow more directional influence to be accounted for in a data set (important to streams) may also allow improved accuracy with a smaller initial sample size.

A further extension of habitat interpolation methods includes the use of interpolated maps with prediction of species occurrence. Stream maps from interpolated data of substrate, depth, and velocity will allow prediction of species occurrence based on existing knowledge of species requirements for these habitat variables. This use of interpolated maps will benefit rare-species management, when habitat loss or habitat-specificity is an important management concern, such as in some benthic fishes or freshwater mussels. Our further research will focus this approach on the Elk River, West Virginia, and on habitat availability and prediction of occurrence of the Elk River crystal darter.

Acknowledgements

This study was funded by the United States Fish and Wildlife Service. We thank P. Mazik, J. S. Kite, G. Staines and R. Utz for comments and criticisms during creation of the
manuscript and during the scientific process. Reference to trade names does not imply
government endorsement of commercial products.

**Literature Cited**

Arend, K. K. 1999. Classification of streams and reaches. Pages 57-74 *in* M. B. Bain and
N. J. Stevenson, editors. Aquatic habitat assessment: common methods.
American Fisheries Society, Bethesda, Maryland.

Association of American Geographers* 90:146-156.


Atkinson, P. M. 1996. Optimal sampling strategies for raster-based geographical


methods*. American Fisheries Society, Bethesda, Maryland.

Dettmers, R. and J. Bart. 1999. A GIS modeling method applied to predicting forest

Ferguson, R.I. 2003. Emergence of abrupt gravel to sand transitions along rivers through

Fisher, W. L. and F. J. Rahel, editors. 2004. *Geographic information systems in
fisheries*. American Fisheries Society, Bethesda, Maryland.


Greenberg, L. A. 1991. Habitat use and feeding behavior of thirteen species of benthic

He, H. S., D. J. Madenoff, T. A. Sickley, and G. G. Guntenspergen. 2000. GIS
interpolations of witness tree records (1839-1866) for northern Wisconsin at


Table 1. Substrate size categories of the United States Department of Agriculture (Soil Survey Division Staff 1993).

<table>
<thead>
<tr>
<th>Substrate Name</th>
<th>Diameter in millimeters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stones</td>
<td>&gt; 250</td>
</tr>
<tr>
<td>Cobble</td>
<td>250–76</td>
</tr>
<tr>
<td>Gravel</td>
<td>76–2</td>
</tr>
<tr>
<td>Very coarse sand</td>
<td>2–1</td>
</tr>
<tr>
<td>Coarse sand</td>
<td>1–0.5</td>
</tr>
<tr>
<td>Medium sand</td>
<td>0.5–0.25</td>
</tr>
<tr>
<td>Fine sand</td>
<td>0.25–0.10</td>
</tr>
<tr>
<td>Very fine sand</td>
<td>0.10–0.005</td>
</tr>
<tr>
<td>Coarse silt</td>
<td>0.005–0.002</td>
</tr>
</tbody>
</table>
Table 2. Substrate root mean square error values of complete substrate based on actual values, 5% interpolation and individual 5% substrate type root mean square error values.

<table>
<thead>
<tr>
<th>RMSe Value</th>
<th>SUBSTRATE</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.074</td>
<td>Gravel</td>
</tr>
<tr>
<td>0.049</td>
<td>Silt</td>
</tr>
<tr>
<td>0.046</td>
<td>Boulder</td>
</tr>
<tr>
<td>0.039</td>
<td>All Substrate (actual)</td>
</tr>
<tr>
<td>0.034</td>
<td>Sand</td>
</tr>
<tr>
<td>0.033</td>
<td>Cobble</td>
</tr>
<tr>
<td>0.030</td>
<td>5% Interpolation</td>
</tr>
<tr>
<td>0.028</td>
<td>Land</td>
</tr>
</tbody>
</table>
Table 3. Depth root mean square error values based on actual values, 5% interpolation, and 2.5% interpolation.

<table>
<thead>
<tr>
<th>RMSe</th>
<th>DEPTH</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.38533</td>
<td>Complete Depth (Actual)</td>
</tr>
<tr>
<td>0.36623</td>
<td>5% Interpolation</td>
</tr>
<tr>
<td>0.34491</td>
<td>2.5% Interpolation</td>
</tr>
</tbody>
</table>
List of Figures

Figure 1. Simple flow chart containing general overview of data collection and interpolation process.

Figure 2. Raster maps of substrate at the Aarons Creek study site for actual substrate values (Actual), and interpolated values based on 5% (A) and 2.5% (B) of actual values. Northwest corner of actual and interpolations A/B = (591540, 4386028).

Figure 3. Raster maps of water depth at the Aarons Creek study site for actual depth values (Actual), and interpolated values based on 5% (A) and 2.5% (B) of actual values. Northwest corner of Actual/A/B = (591540, 4386028).

Figure 4. Two methods of accuracy assessments of 2.5% and 5.0% substrate interpolations; (1) the percent match of interpolated and actual cell values of substrate types and (2) the percent match of interpolated and actual areas of substrate types.

Figure 5. The percent difference between interpolated and actual substrate areas for six substrate categories. Interpolation A is based on 5% of actual substrate cell values and interpolation B is based on 2.5% of actual substrate cell values.

Figure 6. Accuracy assessment of the 5% (A) and 2.5% (B) depth interpolations using the percent of interpolated values within several ranges of true values based on thresholds of 5% (3 cm), 10% (6 cm), and 20% (12 cm) of maximum site depth (60 cm).
Figures

Data collection at site substrate/depth, site characteristics, UTM coordinates

Enter data into Excel

Import site corner coordinates to ArcMap and geo-reference

Export adjusted corners to Excel and add UTM coordinates to all data points for depth and substrate

Import full data set to ArcMap for digital representation of site at 1/3 meter square

Create 2.5% and 5.0% data point subsets, each point selected based on location to edge boundary and feature terminus for interpolation

Interpolate all data sets; complete, 2.5% and 5% and extract values to all initial data points for comparison to initial recorded values (complete data set)

Export all interpolated values and linked complete site values for comparison

Figure 1. Simple flow chart containing general overview of data collection and interpolation process.
Figure 2. Raster maps of substrate at the Aarons Creek study site for actual substrate values (Actual), and interpolated values based on 5% (A) and 2.5% (B) of actual values. Northwest corner of actual and interpolations A/B = (591540, 4386028).
Figure 3. Raster maps of water depth at the Aarons Creek study site for actual depth values (Actual), and interpolated values based on 5% (A) and 2.5% (B) of actual values. Northwest corner of Actual/A/B = (591540, 4386028).
Figure 4. Two methods of accuracy assessments of 2.5% and 5.0% substrate interpolations; (1) the percent match of interpolated and actual cell values of substrate types and (2) the percent match of interpolated and actual areas of substrate types.
Figure 5. The percent difference between interpolated and actual substrate areas for six substrate categories. Interpolation A is based on 5% of actual substrate cell values and interpolation B is based on 2.5% of actual substrate cell values.
Figure 6. Accuracy assessment of the 5% (A) and 2.5% (B) depth interpolations using the percent of interpolated values within several ranges of true values based on thresholds of 5% (3 cm), 10% (6 cm), and 20% (12 cm) of maximum site depth (60 cm).
Appendix

Appendix 1. A PVC grid (0.093 m²) used for substrate sampling.