Restoring Brook Trout to North Carolina’s National Forests:
Using GIS-based methods to evaluate and target watersheds for reintroduction

by

Sara DiBacco
Dr. Dean Urban, Advisor
May 2008

Masters project submitted in partial fulfillment of the requirements for the Master of Environmental Management degree in the Nicholas School of the Environment and Earth Sciences of Duke University

2008
ABSTRACT

Stream systems in the Southern Appalachian Mountains represent the southern limit of brook trout (*Salvelinus fontinalis*) distribution in the eastern United States and are home to the region’s only native salmonid, the Southern Appalachian Brook Trout. Currently, the species occupies only 25% of its former range in the region, but in North Carolina, opportunities exist to restore brook trout to high quality National Forest watersheds. The purpose of this project is to provide the information necessary to design a watershed-specific restoration strategy for the Fires Creek Watershed, Nantahala National Forest and to develop a model that predicts natural migration barriers within a stream network.

A geographic information system (GIS) is the primary analysis tool used to derive, interpret, and display relevant data. In the Fires Creek Watershed, migration barriers are identified and characterized to delineate potential brook trout reintroduction sites. The watershed is also assessed as a target for brook trout restoration according to five criteria. These are the historical presence of brook trout, the current distribution of trout in the basin, the genetic identity of potential donor populations, site accessibility, and current and future habitat suitability. Barrier data are also used to develop a classification and regression tree (CART) model to predict barrier locations.

Results show that numerous opportunities exist to restore brook trout to the Fires Creek Watershed. The most effective restoration strategy combines the availability of protected habitat, as delineated by migration barriers, with information extracted from the comprehensive assessment. The model provides the framework for a tool that improves the efficiency of completing restoration projects, but results suggest that higher resolution data is necessary to increase prediction success. Overall, this work contributes to the development and implementation of brook trout restoration projects in North Carolina’s National Forests.
# Table of Contents

INTRODUCTION .......................................................................................................................................................1

OBJECTIVES................................................................................................................................................................3

STUDY AREA .............................................................................................................................................................4

METHODS...................................................................................................................................................................6
  DATA SOURCES .........................................................................................................................................................6
  I. IDENTIFICATION AND CATEGORIZATION OF TROUT MIGRATION BARRIERS ............................................7
  II. ASSESSMENT OF THE FIRES CREEK WATERSHED AS A TARGET FOR RESTORATION ......................10
  III. DEVELOPMENT OF A MODEL TO PREDICT MIGRATION BARRIERS ..................................................13

RESULTS....................................................................................................................................................................18
  I. IDENTIFICATION AND CATEGORIZATION OF TROUT MIGRATION BARRIERS ........................................18
  II. ASSESSMENT OF THE FIRES CREEK WATERSHED AS A TARGET FOR RESTORATION ..................21
  III. DEVELOPMENT OF A MODEL TO PREDICT MIGRATION BARRIERS ..................................................28

DISCUSSION............................................................................................................................................................35
  IMPLICATIONS FOR BROOK TROUT RESTORATION IN THE FIRES CREEK WATERSHED ....................35
  IMPLICATIONS FOR MODEL DEVELOPMENT AND APPLICABILITY ........................................................38

CONCLUSIONS..........................................................................................................................................................42

ACKNOWLEDGEMENTS ........................................................................................................................................43

REFERENCES ............................................................................................................................................................44

APPENDIX A: GIS DATA SOURCES .....................................................................................................................47

APPENDIX B: ADDITIONAL STREAM HABITAT AND BARRIER NOTES ..................................................48

APPENDIX C: SELECTED GIS MODELS AND SCRIPTS ..................................................................................51

APPENDIX D: SELECTED R CODE FOR CART MODEL ..................................................................................58
LIST OF FIGURES

FIGURE 1. STUDY AREA....................................................................................................................5
FIGURE 2. BARRIER EXAMPLES ......................................................................................................9
FIGURE 3. MAJOR AND MODERATE BARRIERS IN THE FIRES CREEK WATERSHED....................20
FIGURE 4. TEMPERATURE AND pH VALUES RECORDED IN THE FIRES CREEK WATERSHED........23
FIGURE 5. DISTRIBUTION OF STREAM HABITAT AND WATERSHED AREA WITHIN FIVE ELEVATION ZONES. ..............................................................25
FIGURE 6. DISTRIBUTION AND GENETIC IDENTITY OF BROOK TROUT POPULATIONS IN THE HIWASSEE RIVER BASIN. .27
FIGURE 7. CART TREE DIAGRAMS FOR MODEL RUN 1 AND 2............................................................30
FIGURE 8. CONFUSION MATRICES FOR MODEL RUN 1 AND 2 ..........................................................31
FIGURE 9. CLASSIFICATION SUCCESS MEASURES FOR MAPPED RESULTS IN THE FIRES CREEK WATERSHED ........31
FIGURE 10. PREDICTED LOCATIONS OF BARRIERS WITHIN THE FIRES CREEK WATERSHED................32
FIGURE 11. CLASSIFICATION SUCCESS MEASURES FOR MAPPED RESULTS IN COLD SPRING CREEK WATERSHED ....33
FIGURE 12. PREDICTED LOCATIONS IN CHERRY CREEK AND UPPER-COLD SPRING CREEK. ..................34

LIST OF TABLES

TABLE 1. MIGRATION BARRIERS ARE ASSIGNED TO ONE OF THREE CATEGORIES DESCRIBING BARRIER SUITABILITY....8
TABLE 2. VARIABLES CREATED FOR USE IN THE CART MODEL..........................................................17
TABLE 3. BARRIERS IDENTIFIED AS ‘MAJOR’ AND ‘MODERATE’ IN THE FIRES CREEK WATERSHED........19
TABLE 4. STREAM HABITAT WITHIN FIVE ELEVATION ZONES. ................................................................24
TABLE 5. CATCHMENTS WITH DRAINAGE AREAS LESS THAN 3 KILOMETERS AND TOTAL AREA ABOVE 1,067M. ........26
TABLE 6. MINIMUM DISTANCE FROM EACH BARRIER TO A ROAD AND TRAIL. ............................................26
TABLE 7. COMPARISON OF PREDICTOR VARIABLES FOR FIRES CREEK AND COLDSPRING CREEK WATERSHEDS........33
**INTRODUCTION**

Stream systems in the Southern Appalachian Mountains are home to the region’s only native salmonid species, the Southern Appalachian Brook Trout (SABT). SABT are a genetically unique species of brook trout (*Salvelinus fontinalis*), distinct from those occupying ranges in the northeastern United States and Canada. Historically, the species occupied over 9,600 kilometers of stream habitat from the New River drainage in Virginia to the southern extent of the Appalachian Mountains in Georgia. Currently, SABT occupy only 25% of their native range in the Southern Appalachians (Habera and Moore 2005).

In the early 1900s, logging and land clearing practices severely degraded stream habitat and nearly eradicated SABT from the region. As land use regimes shifted and watersheds recovered, management agencies began stocking streams with rainbow (*Oncorhynchus mykiss*) and brown (*Salmo trutta*) trout from the western United States and Europe, respectively (Larson and Moore 1985; Habera and Moore 2005). However, these efforts to replenish stream fish populations impaired the recovery of native brook trout (Habera and Moore 2005).

The introduction of non-native salmonids, particularly rainbow trout, decreased brook trout distribution in lower elevation and lower latitude streams (Larson and Moore 1985; Flebbe 1997). The ability of rainbow trout to displace brook trout was partly attributed to their larger average size, greater fecundity, and fewer year-0 class failures (Larson and Moore 1985; Clark and Rose 1997). Although recent evidence suggests that competition with non-native trout has stabilized, remaining brook trout populations are mostly restricted to smaller, high elevation headwater streams (Flebbe 1997; Habera et al. 2001). These habitat areas are susceptible to extreme weather events, such as flooding and drought, and are also characterized by low productivity (Larson and Moore 1985).
In addition to rainbow and brown trout, streams were stocked with hatchery-derived brook trout from ranges in the northeastern United States (Habera et al. 2001). Genetic tests performed in the 1990s revealed that brook trout populations native to the Southern Appalachians were “taxonomically distinct” from hatchery-derived northern brook trout (McCracken et al. 1993; Hayes et al. 1996). Results also demonstrated that hybridization was occurring between SABT and northern brook trout, thus threatening their unique genetic heritage (Hayes et al. 1996). In North Carolina, sampling in 29% of streams known to support wild brook trout revealed that only 62% contain genetically distinct SABT populations (Habera and Moore 2005).

SABT also continue to face threats from range contraction, habitat fragmentation, and environmental change (Larson and Moore 1985; Hayes et al. 1996; Flebbe 2006). Yet, with an awareness of these persistent threats, opportunities exist to ensure the long-term viability of the region’s only native trout species. The Forest Service in North Carolina is committed to restoring SABT to high quality National Forest watersheds (S. Bryan, United States Forest Service, pers. comm.). The restoration process involves identification of viable trout migration barriers in the watershed, validation of available fish distribution information, and the physical reintroduction of brook trout at selected sites.

The identification of migration barriers is necessary to improve the viability of reintroduced brook trout by preventing sympatric associations with non-native trout (Larson and Moore 1985; Habera and Moore 2005). Thus, by restricting the upstream movement of fish, migration barriers delineate protected habitat areas and indicate potential reintroduction sites (Williams 1996; Thompson and Rahel 2003). In addition, the location of barriers within
candidate stream networks provides a framework for further evaluating the suitability of a watershed for brook trout restoration.

**OBJECTIVES**

The overall goal of this Master’s Project is to support ongoing efforts to restore the native and genetically distinct Southern Appalachian Brook Trout to watersheds throughout North Carolina’s National Forests. Work on this project was initiated during an internship with the Southern Appalachian Forest Coalition (SAFC) and in collaboration with the United States Forest Service (USFS). The specific objectives of this Master’s Project are as follows:

I. Identify and categorize trout migration barriers in the Fires Creek Watershed.

II. Perform a comprehensive assessment of the Fires Creek Watershed as a target for restoration.

III. Develop a classification and regression tree model (CART) to predict the location of natural migration barriers within a stream network.

Objectives I and II provide the information necessary to create a watershed-specific brook trout restoration strategy for Fires Creek. The assessment also serves as a framework for evaluating the suitability of other candidate watersheds. Objective III improves the overall efficiency of completing brook trout restoration projects by providing a tool that accomplishes the first step; identifying the location of natural barriers. The model is developed from widely available data and popular software to facilitate its application in other candidate watersheds.
STUDY AREA

Communication with members of the Forest Service indicates that the Fires Creek Watershed is a high priority site for brook trout restoration (S. Bryan, United States Forest Service, *pers. comm.*). The watershed is located along the northern border of Clay County, NC between the Valley River and Tusquitee mountain ranges (Figure 1). It drains approximately 77 square kilometers and belongs to one of two major sub-basins comprising the Hiwassee River Basin. The watershed is identified by the unique 14-digit hydrologic unit code, 06020002071010 (USDA-NRCS 1995).

The 2005 Hiwassee Basin Report describes the area as a mostly undisturbed forested catchment characterized by high water quality and stream habitat (North Carolina Division of Environment and Natural Resources - NCDENR-DWQ). Fires Creek is classified as an Outstanding Water Resource and as Wild Trout Waters by NCDENR and the NC Wildlife Resource Commission (NCWRC), respectively. Fires Creek and its watershed are designated State Natural Heritage Areas due to the presence of rare and endangered aquatic and riparian organisms (NCDENR-DPR).

A 3 kilometer portion of Fires Creek located above State Road 1334 is annually stocked between March and June with over 2000 rainbow, brown, and brook trout (NCDENR-DWQ). Wild, young of the year rainbow trout have been collected from reaches upstream of the stocking site between 1999 and 2004 (NCDENR-DWQ). Fish distribution surveys provided by members of North Carolina’s National Forest Service show that naturalized populations of rainbow trout occur throughout the main Fires Creek channel. Wild brook trout are not documented in these records, which extend back to 1978 (USFS-National Forests in NC).
The main study area drains approximately 55 square kilometers, including the headwaters originating in the Valley River and Tusquitee Mountains (Figure 1). The area falls completely within the Nantahala National Forest except for a 0.2 square kilometer private in-holding at the top of Laurel Creek. Streams in the study area occupy over 57 kilometers and span an elevation of 492 – 1605 meters. Roads and trails owned and maintained by the Forest Service provide access throughout much of the watershed. However, assessments performed by SAFC in 2006 identified numerous locations in which the road’s proximity to the creek was a potential source of sedimentation. Culvert and sediment fence failings were also noted.

**Figure 1.** The overview map shows the location of Clay County in North Carolina and the lower map demonstrates the watershed area and streams within the study site.
METHODS

Data Sources

Digital fish survey data for the Hiwassee River Basin and digital road and trail layers for
the Nantahala National Forest were provided by members of the North Carolina National Forest
Service. Genetic information for brook trout populations in the Hiwassee River Basin was
provided by the North Carolina Wildlife Resources Commission. LiDAR derived digital
elevation models for Clay and Haywood counties in North Carolina were downloaded from the
North Carolina Department of Transportation (http://www.ncdot.org/it/gis/DataDistribution/
ContourElevation). The National Hydrography Dataset Plus (NHD Plus) for hydrologic region 6
was provided by the United States Environmental Protection Agency and the United States
Geological Survey (USGS) (http://www.horizon-systems.com/nhdplus/data.php). Soil data was
accessed from the Soil Data Mart supported by the Natural Resources Conservation Service
(http://soildatamart.nrcs.usda.gov). National land cover data from 2001 including percent
canopy cover was obtained from USGS (http://gisdata.usgs.net/website/MRLC) (Also see GIS
Data Sources in Appendix A).
I. Identification and Categorization of Trout Migration Barriers

Potential trout migration barriers in the Fires Creek watershed were identified through stream surveys. Surveys consisted of walking the streams and documenting the location of candidate structures using a handheld global positioning system (GPS) (Garmin GPS Map 76S). Potential barriers were photographed and where applicable, physical measurements describing fall height, cascade length, slope, total drop, and total length were obtained. The presence or absence of potential resting pools was also documented. Structures that consisted of multiple feature types were referenced as barrier complexes, e.g. a waterfall followed by a cascade. In addition to measurements of each individual ‘step’ within a complex, the sequence of features was noted.

Because the determination of exactly what constitutes an effective barrier to upstream trout migration is subject to uncertainty, expert advice and information from literature was sought. Sheryl Bryan, the Fisheries Biologist for North Carolina’s National Forest, reviewed photographs from a preliminary survey of barriers and indicated which feature types and characteristics served as effective barriers. Jim Herrig, the Aquatic Biologist for Cherokee National Forest, advised that man-made structures used as trout barriers are constructed to be 6 - 8 feet in height to prevent trout passage by jumping and also avoidance during flood conditions. In a paper specifically describing brook trout jumping ability, Kondraieff and Myrick (2006) demonstrated that the maximum jump height of the largest brook trout (> 8 inches) considered in the study was 29 inches with a preceding plunge pool depth of at least 16 inches. Based on this information, the physical measurements and observations made in the field were used to qualitatively rank each potential barrier as minor, moderate, or major (Table 1, Figure2).
Table 1. Migration barriers are assigned to one of three categories describing barrier suitability.

<table>
<thead>
<tr>
<th><strong>Category</strong></th>
<th><strong>Description</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Minor</strong></td>
<td>Structures are likely not significant barriers to upstream passage. They have shear waterfall heights less than or equal to 30 inches and/or cascade slopes less than 20 degrees.</td>
</tr>
<tr>
<td><strong>Moderate</strong></td>
<td>Features are potentially more or less severe depending on flow conditions. They have a shear waterfall height greater than 30 inches, and/or a total drop greater than 44 inches with slope at least 30 degrees, and/or a cascade length greater than 4 feet with slope at least 45 degrees.</td>
</tr>
<tr>
<td><strong>Major</strong></td>
<td>Structures are completely impassable under any plausible flow conditions due to shear waterfall height, and/or the combination of cascade length and slope, and/or the presence of impassable road culverts.</td>
</tr>
</tbody>
</table>
Figure 2. The pictures demonstrate examples of barriers categorized as minor, moderate and major.
II. Assessment of the Fires Creek Watershed as a Target for Restoration

Multiple criteria were used to evaluate the suitability of the Fires Creek Watershed as a target for brook trout reintroduction and to further inform the design of a watershed specific restoration strategy. Several restoration guidelines suggested by the Southern Division of the American Fisheries Society’s Trout Committee were adapted for use in this assessment (Habera and Moore 2005), in addition to habitat parameters specified by other researchers (Williams 1996; Petty et al. 2005). Logistical considerations such as site accessibility and the location of potential donor populations were also addressed to assess project feasibility. A geographic information system (GIS) was the primary tool used to characterize potential habitat in the Fires Creek Watershed and to extract information relevant to the design of a successful brook trout restoration strategy (ArcGIS version 9.2, ESRI, Redlands, CA).

Fish Distribution:

The historical presence of brook trout indicates whether a watershed falls within the species’ native range. In the absence of precise data, the existence of reproducing populations of non-native, cold water trout species can serve as evidence that the site is within the native range of brook trout (Habera and Moore 2005; Hudy et al. 2005). The current distribution of brook trout in the Hiwassee River Basin identifies potential donor populations for reintroduction. In addition, the present distribution of introduced rainbow and brown trout informs the scope of restoration activities necessary to restore brook trout in the watershed.

Fish distribution information was used to confirm the historical presence of brook trout in the watershed, to locate potential donor populations, and to demonstrate the extent of habitat occupied by non-native trout. Spatially referenced fish distribution data were provided by the
Forest Service and displayed in a GIS. The data layer was overlain on 14-digit hydrologic units to identify their location within the Hiwassee River Basin.

Genetics:

The genetic identity of potential donor populations is required to accomplish restoration of the genetically distinct Southern Appalachian Brook Trout (Habera and Moore 2005). Results of genetic analyses were matched to spatially referenced brook trout populations and displayed to identify the closest population with the appropriate lineage.

Site Accessibility:

The location of roads and trails in the watershed provide an indication of the physical work necessary to reintroduce fish at certain locations. The distance to the nearest road and nearest trail was calculated from each barrier location to depict a minimum travel distance to habitat located above that site. Roads and trails were differentiated to suggest the proper mode of transportation.

Habitat:

The quality and quantity of brook trout habitat are important determinants of brook trout population success. Specifically, stream temperature and pH affect many aspects of in-stream habitat quality. Brook trout are a coldwater fish species that prefer temperatures less than 20ºC, with an optimal range for growth and survival between 11 and 16ºC (Swift and Messer 1971; Raleigh 1982). They can tolerate pH values between 4.0 and 9.5, but a range of 6.8 – 8.0 is optimal (Raleigh 1982). As measured by stream temperature and pH, habitat suitability is potentially impacted by future environmental changes. In the Southeast, streams below 915m are susceptible to warming associated with global climate change, and likewise, streams above 1,067m are particularly sensitive to acidification from natural and anthropogenic sources (Habera
and Moore 2005). Tributaries with basin areas less than 3 square kilometers are identified as preferred spawning sites for brook trout (Petty et al. 2005). Researchers have also indicated a minimum drainage area of approximately 2.5 square kilometers above an elevation of 1,067 meters as adequate habitat (Williams 1996).

Current stream temperatures and pH values within the watershed were surveyed using a handheld electronic pH and temperature device (YSI® EcoSense® pH 10 pH/Temperature Pen, Forestry Suppliers). The values and corresponding GPS locations were mapped in a GIS to demonstrate ‘snapshots’ of temperature and pH conditions within the study area. Habitat areas that are potentially sensitive to environmental changes were quantified by tabulating the area of the stream raster within each of five elevation zones (492 – 610, 611 – 762, 763 – 914, 915 – 1066, and 1067 – 1605 meters). The same procedure was used to calculate the percentages of individual catchment areas falling within each elevation zone. The amount of stream habitat with catchment areas less than 3 square kilometers and the number of drainage areas above an elevation of 1,067m were calculated by combining flow line attribute information with elevation data in a GIS.
III. Development of a Model to Predict Migration Barriers

A classification and regression tree model (CART) was developed to predict the occurrence of natural migration barriers within a stream network.

CART:

CART models are flexible tools frequently used in ecology to model habitat and species distributions (Breiman et al. 1984; De’ath and Fabricius 2000; Vayssieres et al. 2000; Urban 2002). The model is discretely applied as either a classification or regression tree depending on whether the response variable is categorical or continuous. In both cases, the result is a hierarchical ‘tree’ diagram depicting the combination and sequence of predictor variables that indicate a specific response (Urban 2002).

CART models use an algorithm that recursively partitions data into subsets of a response group according to a set of predictor variables (Vayssieres et al. 2000; Urban 2002). Each partition maximizes the homogeneity within subsets of the response group. The first branch of the tree is generated by the predictor variable and corresponding rank value that achieves the best separation of response group subsets. The process is repeated for each subsequent branch until the responses are split into pure groups or a user-defined threshold is reached.

Classification trees are designed to minimize misclassification of the data used in model development. However, the model is generally over-fitted and not suitable for classifying new data. To improve robustness, a cross-validation method is used. This procedure recreates the model \((n\) times as specified by the user) with \((n-1)/n\) of the original input data and uses the remaining \(1-((n-1)/n)\) to assess misclassification rates. For binary responses, cross-validation produces a confusion matrix from which accuracy measures are calculated (Vayssieres et al. 2000; Urban 2002).
Several features of CART models make them useful for ecological applications and convenient for the purposes of this project. Often, patterns and processes observed in the natural world are a result of complex biological, physical, and chemical relationships. Such complexity is not adequately explained by linear methods that require data to conform to particular distributional assumptions. As a nonparametric approach, CART models do not require explanatory data to assume any distributional forms. In addition, CART methods avoid some disadvantages associated with logistic regression by relying on an algorithm to reveal and distinguish multiple unique relationships present within the data (Vayssieres et al. 2000; Urban 2002). Finally, model outputs are easily mapped in a GIS as a series of conditional statements that lead to response group membership (Urban 2002). The advantages of CART methods make them well suited to deal with the highly variable structures that comprise natural migration barriers and the non-homogenous landscapes characteristic of high elevation, headwater stream systems.

*Model Development*

A GIS was used to generate a response variable layer and several explanatory variable datasets for the Fires Creek Watershed (Table 2). The response variable consisted of the stream raster layer in which all cells corresponding to locations of ‘Major’ and ‘Moderate’ barriers were coded as 1s. All non-barrier locations were coded as 0s. The result was a response dataset that contained 11,973 zeros and 26 ones.

Predictor variables were derived from landscape characteristics that potentially indicate the occurrence of natural barriers within a stream. Datasets were created for slope, local relief, stream gradient, percent canopy cover, and soil factors (Table 2). Slope, local relief, and stream gradient variables describe the topography surrounding and underlying the stream network.
Specifically, local relief variables were designed to accentuate variations in elevation in the stream and riparian areas. They were generated by performing 3x3 or 5x5 neighborhood, or focal, manipulations of the digital elevation model. Percent canopy cover provided a measure of openness surrounding a stream, as occurs near some ‘Major’ features and indicated the potential for large woody debris contributions. Soil characteristics indicated a susceptibility to sediment, boulder, tree, and other debris inputs from riparian areas. Overall, the predictor variables were intended to capture aspects of non-homogeneity in the composition of barriers and in the stream and riparian landscape.

The coded stream raster was used to sample values from all predictor variables that overlapped with barrier and non-barrier (i.e. where cell value = 1 or 0) locations in the streams. Each row corresponded to a barrier or non-barrier cell in the stream raster and contained values for each predictor variable. Since there were an overwhelming number of non-barrier cells in the data, a subset of 100 absence rows was randomly selected and re-joined to the presence only rows. The final dataset contained 100 absence rows and 26 presence rows.

The library package “tree” (Ripley 2006) contains the commands and functions necessary to run the CART algorithm in R (R Development Core Team 2007). Initially, the model was specified to include all predictor variables, but soil attributes were later removed (Table 2). A classification tree was generated that separates barrier and non-barrier responses according to specific values and sequences of selected predictor variables.

The predict function was used to produce a confusion matrix that tabulates responses predicted by the fitted model against the actual, or true, responses present in the input data. Measures of classification success were derived from the matrix by tallying the number of true positive, true negative, false positive and false negative matches. After cross-validation and tree-
pruning procedures, the predict process was repeated for a tree pruned to an optimal number of branches.

Interaction pathways that led to barrier responses in the model tree were interpreted as a series of conditional statements. These conditions, imposed by the sequence and value of predictor variables, were nested in a map algebra command in a GIS. The output was a raster layer which indicated the value of 1 in all cells that met the specified conditions. Mapping conditional statements in the Fires Creek Watershed allowed for comparison of predicted and known barrier locations. The conditional statements were also applied to the Cold Spring Creek Watershed using watershed-specific environmental data to predict and map barrier locations. Predicted locations in two streams, Cherry Creek and the upper section of Cold Spring Creek, were validated using barrier data collected in March 2008.
Table 2. The table demonstrates the variables created for use in the CART model.

<table>
<thead>
<tr>
<th><strong>Response:</strong></th>
<th><strong>Data</strong></th>
<th><strong>Description</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>1 or 0</td>
<td>Raster dataset of NHD Plus flowlines</td>
<td>Raster dataset used to sample values from predictor variable datasets. Cells coded as one indicate barrier locations (n=26). All other cells are zero (n=11,973).</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Predictors:</strong></th>
<th><strong>Data</strong></th>
<th><strong>Description</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Slope</td>
<td>DEM (LiDAR - 20ft) derived slope</td>
<td>Slope represents the maximum change in elevation over the distance between a cell and its eight neighbors or a 3x3 focal window. High slopes demonstrate areas of abrupt terrain change.</td>
</tr>
<tr>
<td>Local relief</td>
<td>DEM derived 5x5 focal mean minus DEM</td>
<td>Focal manipulation of the DEM accentuates elevation variations. Relative elevation depressions, or coves, and relative elevation peaks, or knobs, are shown by subtracting actual elevation values from a focal window mean. Large positive values indicate coves and large negative values indicate knobs. Barrier locations are potentially indicated by a variable that captures these aspects of non-homogeneity in the landscape. The smaller focal window size (3600 square feet or 334 square meters) captures terrain changes within a more immediate distance to the streams, which are relatively narrow in width.</td>
</tr>
<tr>
<td>Local relief</td>
<td>DEM derived 3x3 focal mean minus DEM</td>
<td>Focal manipulation of the DEM accentuates elevation variations. Relative elevation depressions, or coves, and relative elevation peaks, or knobs, are shown by subtracting actual elevation values from a focal window mean. Large positive values indicate coves and large negative values indicate knobs. Barrier locations are potentially indicated by a variable that captures these aspects of non-homogeneity in the landscape. The smaller focal window size (3600 square feet or 334 square meters) captures terrain changes within a more immediate distance to the streams, which are relatively narrow in width.</td>
</tr>
<tr>
<td>Local relief</td>
<td>DEM derived 5x5 focal max - 5x5 focal min</td>
<td>Focal manipulation of the DEM accentuates elevation variations. Areas in which sharp elevation changes occur are indicated by subtracting the focal minimum elevation value from the focal maximum elevation value. Large values indicate areas of high elevation contrast. Smaller numbers depict areas which have relatively small changes in elevation. Both focal window sizes are reasonable but the larger window (10,000 square feet or 929 square meters) captures elevation change over a larger area surrounding the stream; however, high values derived from a smaller focal window manipulation indicates a more localized, sharp change in terrain near the stream.</td>
</tr>
<tr>
<td>Local relief</td>
<td>DEM derived 3x3 focal max - 3x3 focal min</td>
<td>Focal manipulation of the DEM accentuates elevation variations. Areas in which sharp elevation changes occur are indicated by subtracting the focal minimum elevation value from the focal maximum elevation value. Large values indicate areas of high elevation contrast. Smaller numbers depict areas which have relatively small changes in elevation. Both focal window sizes are reasonable but the larger window (10,000 square feet or 929 square meters) captures elevation change over a larger area surrounding the stream; however, high values derived from a smaller focal window manipulation indicates a more localized, sharp change in terrain near the stream.</td>
</tr>
<tr>
<td>Stream gradient</td>
<td>Slope of flowline from NHD Plus attributes</td>
<td>Stream slope is generated by subtracting the minimum and maximum elevation values derived for each NHD Plus flowline and then dividing by flowline length. It provides a measure of the overall steepness associated with each flowline.</td>
</tr>
<tr>
<td>Canopy cover</td>
<td>3x3 focal mean of percent canopy cover from 2001 NLCD</td>
<td>Percent canopy cover is a proxy for tree density which indicates a potential for large woody debris contributions and demonstrates a measure of openness in riparian areas which are associated with some 'Major' barrier features.</td>
</tr>
<tr>
<td>Soils</td>
<td>Erodibility factor Kf</td>
<td>Soil characteristics indicate points along the stream that are susceptible to bank collapse, boulder deposition, and fallen trees. Values for the erodibility factor and percent clay are derived by equalizing the percent map unit depicting major soils to 100% and then calculating a weighted average of the value indicated for the erodibility factor, Kf, or % clay.</td>
</tr>
<tr>
<td>Soils</td>
<td>Percent clay</td>
<td>Soil characteristics indicate points along the stream that are susceptible to bank collapse, boulder deposition, and fallen trees. Values for the erodibility factor and percent clay are derived by equalizing the percent map unit depicting major soils to 100% and then calculating a weighted average of the value indicated for the erodibility factor, Kf, or % clay.</td>
</tr>
</tbody>
</table>
RESULTS

I. Identification and Categorization of Trout Migration Barriers

In total, 19 ‘Moderate’ and 8 ‘Major’ barriers are identified in a thorough field survey for trout migration barriers in the Fires Creek Watershed (Table 3, Figure 2). A road culvert, which represents a man-made migration barrier, is included in the ‘Major’ category but not considered during model development (Barrier ID: 340C+CCB). ‘Minor’ structures are eliminated as potential trout migration barriers and therefore not documented in the field. As previously described, ‘Major’ barriers are considered impassable under any plausible flow conditions given the size and breadth of the structures (Table 1). The impassability of barriers categorized as ‘Moderate’ may vary depending on the precise features of individual structures and the current flow conditions.

Overall, stream surveys in the Fires Creek Watershed demonstrate that numerous opportunities exist to reintroduce brook trout above ‘Major’ and ‘Moderate’ barriers. Much of the habitat within the study area has been explored; however, conditions above the following barriers were not observed: LWD FALL, TC1, and BSB1 (See Additional Notes in Appendix B).
Table 3. The table lists barriers identified as ‘Major’ and ‘Moderate’ in the Fires Creek Watershed with corresponding stream and reach characteristics.

<table>
<thead>
<tr>
<th>Streams</th>
<th>Flowline &amp; Catchment ID</th>
<th>Flowline Length (km)</th>
<th>Drainage Area (km²)</th>
<th>Elevation Range (m)</th>
<th>Barrier ID</th>
<th>Barrier Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fires Creek</td>
<td>19674 801</td>
<td>6.1</td>
<td>7.8</td>
<td>876.4 - 1421.9</td>
<td>U-FC4</td>
<td>MODERATE</td>
</tr>
<tr>
<td></td>
<td>U-FC1</td>
<td>MODERATE</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>U-FC6</td>
<td>MODERATE</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>U-L FC6</td>
<td>MODERATE</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>U-L FC5</td>
<td>MODERATE</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>FC2</td>
<td>MODERATE</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>FC1</td>
<td>MODERATE</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>19674 233</td>
<td>1.3</td>
<td>10.4</td>
<td>826.7 - 876.4</td>
<td>L-LBC2</td>
<td>MAJOR</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>L-LBC10</td>
<td>MAJOR</td>
</tr>
<tr>
<td></td>
<td>Long Branch</td>
<td>L-LBC9</td>
<td></td>
<td></td>
<td>L-LBC7</td>
<td>MODERATE</td>
</tr>
<tr>
<td></td>
<td></td>
<td>L-LBC6</td>
<td></td>
<td></td>
<td>L-LBC6</td>
<td>MODERATE</td>
</tr>
<tr>
<td></td>
<td></td>
<td>L-LBC4</td>
<td></td>
<td></td>
<td>L-LBC4</td>
<td>MAJOR</td>
</tr>
<tr>
<td></td>
<td></td>
<td>L-LBC2</td>
<td></td>
<td></td>
<td>L-LBC2</td>
<td>MODERATE</td>
</tr>
<tr>
<td></td>
<td>19674 199</td>
<td>5.0</td>
<td>4.4</td>
<td>928.0 - 1368.3</td>
<td>LBC2</td>
<td>MAJOR</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>LBC1</td>
<td>MODERATE</td>
</tr>
<tr>
<td></td>
<td>Rockhouse Creek</td>
<td>19674 241</td>
<td>3.0</td>
<td>2.6</td>
<td>679.6 - 1154.0</td>
<td>RC1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>RC5</td>
<td>MODERATE</td>
</tr>
<tr>
<td></td>
<td></td>
<td>L-RC2</td>
<td></td>
<td></td>
<td>L-RC1</td>
<td>MODERATE</td>
</tr>
<tr>
<td></td>
<td>Laurel Creek</td>
<td>19674 237</td>
<td>3.4</td>
<td>4.3</td>
<td>728.9 - 1121.7</td>
<td>LC1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>LC2</td>
<td>MAJOR</td>
</tr>
<tr>
<td></td>
<td>Little Fires Creek</td>
<td>19674 251</td>
<td>4.7</td>
<td>4.5</td>
<td>721.6 - 1256.2</td>
<td>U-LFC1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>LFC1</td>
<td>MODERATE</td>
</tr>
<tr>
<td></td>
<td>Baldspring Branch</td>
<td>19674 221</td>
<td>3.1</td>
<td>1.5</td>
<td>876.4 - 1393.0</td>
<td>BSB1</td>
</tr>
<tr>
<td></td>
<td>Collett Camp Branch</td>
<td>19674 201</td>
<td>2.9</td>
<td>2.6</td>
<td>928.0 - 1250.9</td>
<td>340C+CCB</td>
</tr>
<tr>
<td></td>
<td>Tom Cove Branch</td>
<td>19674 271</td>
<td>2.0</td>
<td>2.6</td>
<td>578.4 - 1062.6</td>
<td>TC1</td>
</tr>
<tr>
<td></td>
<td>Leatherwood Branch</td>
<td>19674 291</td>
<td>2.5</td>
<td>2.3</td>
<td>590.6 - 1048.6</td>
<td>LWD FALL</td>
</tr>
</tbody>
</table>
Figure 3. The map illustrates the location of ‘Major’ and ‘Moderate’ barriers throughout the stream network. Select barrier IDs are shown.
II. Assessment of the Fires Creek Watershed as a Target for Restoration

Comprehensive analyses are performed to assess the suitability of the Fires Creek watershed as a target for restoration and to inform specific restoration considerations. Fish distribution data provided by members of the United States Forest Service demonstrate that naturally reproducing rainbow trout are present throughout the watershed. Although populations of brook trout are not documented, the presence of rainbow trout serves as evidence that the watershed falls within the native range of brook trout. Furthermore, the persistence of a closely related species indicates that available habitat is conducive to brook trout population success.

Temperature measurements recorded throughout the watershed in early September, 2007 demonstrate values of 15.2°C in the headwaters to 20.6°C in the lowest elevations of the study area. pH values in the range of 5.6 – 6.2 are also recorded (Figure 4). These data are not reported to represent seasonal or long-term trends; instead, they provide a cursory indication of habitat suitability in the Fires Creek Watershed. Long-term data reported in the 2005 Hiwassee Basin report indicate an average annual temperature in Fires Creek of 18.9°C and a pH of 6 (NCDENR-DWQ).

The suitability of brook trout habitat in the Fires Creek Watershed is potentially modified by global climate change and acidification (Fiss and Carline 1993; Habera and Moore 2005; Flebbe et al. 2006). In the event of global warming, streams below 915m are susceptible to increased stream temperatures. This corresponds to 30km or 52% of the total available stream habitat. Conversely, habitat that is potentially affected by acidification, i.e. above 1,067m, represents only 13km or 22% of total available habitat. However, if both impacts are simultaneously considered, only 15 kilometers or 26% of total available habitat remains viable (Table 4, Figure 5).
Habitat is also evaluated according to specific life cycle requirements, such as spawning, and to broadly address habitat heterogeneity. Results demonstrate that 12 stream reaches within the study area are characterized with drainages less than 3 square kilometers (Table 5). The percent of these catchments located above 1,067m varies from 0 – 91% (Table 5).

Potential donor populations are located by linking genetic information with spatial fish distribution data. The map shows that the closest confirmed SABT population resides in a neighboring 14-digit hydrologic unit, 06020002070010 (Figure 6) (USDA-NRCS 1995). In total, four confirmed populations of SABT exist in the Hiwassee River Basin. Three of the four confirmed populations are within the same sub-basin as Fires Creek.

The feasibility of reintroducing brook trout to selected sites within the watershed depends on the proximity of roads and trails. The distance from each barrier to the nearest road or trail represents the most conservative estimate of the travel distance required to reach selected habitat. Calculations show that all barriers are within 60 meters of a road but the distance to trails varies from less than 1 meter to over 1,200 meters (Table 6).
Figure 4. The map demonstrates temperature and pH values recorded in the Fires Creek Watershed in September, 2007.
Table 4. The table demonstrates stream habitat quantified by length and percent of total length within five elevation zones.

<table>
<thead>
<tr>
<th>Elevation Categories</th>
<th>Feet (ft)</th>
<th>Meters (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1614-2000</td>
<td>492-610</td>
</tr>
<tr>
<td>2</td>
<td>&gt;2000-2500</td>
<td>611-762</td>
</tr>
<tr>
<td>3</td>
<td>&gt;2500-3000</td>
<td>763-914</td>
</tr>
<tr>
<td>4</td>
<td>&gt;3000-3500</td>
<td>915-1066</td>
</tr>
<tr>
<td>5</td>
<td>&gt;3500-5267</td>
<td>1067-1605</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Stream Reaches in the Study Area (Bold indicates reaches with barriers)</th>
<th>Reach Length (km) in Elevation Categories</th>
<th>Reach Length (%) in Elevation Categories</th>
</tr>
</thead>
<tbody>
<tr>
<td>ID</td>
<td>Stream</td>
<td>Length (km)</td>
</tr>
<tr>
<td>----</td>
<td>----------------------</td>
<td>-------------</td>
</tr>
<tr>
<td>19674221</td>
<td>Baldspring Branch</td>
<td>3.1</td>
</tr>
<tr>
<td>19674205</td>
<td>Coldspring Branch</td>
<td>2.1</td>
</tr>
<tr>
<td>19674217</td>
<td>Coldspring Branch</td>
<td>0.4</td>
</tr>
<tr>
<td>19674201</td>
<td>Collett Camp Branch</td>
<td>2.9</td>
</tr>
<tr>
<td>19674233</td>
<td>Fires Creek</td>
<td>1.3</td>
</tr>
<tr>
<td>19674249</td>
<td>Fires Creek</td>
<td>3.0</td>
</tr>
<tr>
<td>19674269</td>
<td>Fires Creek</td>
<td>4.6</td>
</tr>
<tr>
<td>19674275</td>
<td>Fires Creek</td>
<td>2.2</td>
</tr>
<tr>
<td>19674295</td>
<td>Fires Creek</td>
<td>1.4</td>
</tr>
<tr>
<td>19674339</td>
<td>Fires Creek</td>
<td>1.6</td>
</tr>
<tr>
<td>19674801</td>
<td>Fires Creek</td>
<td>6.1</td>
</tr>
<tr>
<td>19674235</td>
<td>Hickory Cove Creek</td>
<td>2.6</td>
</tr>
<tr>
<td>19674237</td>
<td>Laurel Creek</td>
<td>3.4</td>
</tr>
<tr>
<td>19674243</td>
<td>Laurel Creek</td>
<td>0.5</td>
</tr>
<tr>
<td>19674291</td>
<td>Leatherwood Branch</td>
<td>2.5</td>
</tr>
<tr>
<td>19674293</td>
<td>Leatherwood Branch</td>
<td>0.1</td>
</tr>
<tr>
<td>19674251</td>
<td>Little Fires Creek</td>
<td>4.7</td>
</tr>
<tr>
<td>19674199</td>
<td>Long Branch</td>
<td>5.0</td>
</tr>
<tr>
<td>19674215</td>
<td>Long Branch</td>
<td>1.5</td>
</tr>
<tr>
<td>19674223</td>
<td>Long Branch</td>
<td>0.4</td>
</tr>
<tr>
<td>19674241</td>
<td>Rockhouse Creek</td>
<td>3.0</td>
</tr>
<tr>
<td>19674255</td>
<td>Rockhouse Creek</td>
<td>1.1</td>
</tr>
<tr>
<td>19674203</td>
<td>Tatham Cabin Branch</td>
<td>2.1</td>
</tr>
<tr>
<td>19674271</td>
<td>Tom Cove Branch</td>
<td>2.0</td>
</tr>
<tr>
<td>TOTAL</td>
<td></td>
<td>57.5</td>
</tr>
</tbody>
</table>
Figure 5. The map shows the distribution of stream habitat and watershed area within five elevation zones.
Table 5. The table demonstrates the catchments with drainage areas less than 3 kilometers and the percent of total area above 1,067m.

<table>
<thead>
<tr>
<th>Flowline &amp; Catchment ID</th>
<th>Stream</th>
<th>Drainage Area (km²)</th>
<th>% Above 1,067m (Elevation Category 5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>19674201</td>
<td>Collett Camp Branch</td>
<td>2.5</td>
<td>88.9%</td>
</tr>
<tr>
<td>19674203</td>
<td>Tatham Cabin Branch</td>
<td>1.3</td>
<td>79.9%</td>
</tr>
<tr>
<td>19674205</td>
<td>Coldspring Branch</td>
<td>1.3</td>
<td>90.9%</td>
</tr>
<tr>
<td>19674221</td>
<td>Baldspring Branch</td>
<td>1.5</td>
<td>86.2%</td>
</tr>
<tr>
<td>19674233</td>
<td>Fires Creek</td>
<td>1.3</td>
<td>0.0%</td>
</tr>
<tr>
<td>19674235</td>
<td>Hickory Cove Creek</td>
<td>2.5</td>
<td>36.8%</td>
</tr>
<tr>
<td>19674241</td>
<td>Rockhouse Creek</td>
<td>2.6</td>
<td>41.8%</td>
</tr>
<tr>
<td>19674255</td>
<td>Rockhouse Creek</td>
<td>1.1</td>
<td>0.0%</td>
</tr>
<tr>
<td>19674271</td>
<td>Tom Cove Branch</td>
<td>2.6</td>
<td>17.9%</td>
</tr>
<tr>
<td>19674275</td>
<td>Fires Creek</td>
<td>1.7</td>
<td>0.0%</td>
</tr>
<tr>
<td>19674291</td>
<td>Leatherwood Branch</td>
<td>2.2</td>
<td>15.4%</td>
</tr>
<tr>
<td>19674339</td>
<td>Fires Creek</td>
<td>1.2</td>
<td>0.0%</td>
</tr>
</tbody>
</table>

* Bold indicates barriers are present on the corresponding reach.

Table 6. The table shows the minimum distance from each barrier to a road and trail.

<table>
<thead>
<tr>
<th>Streams</th>
<th>Barrier ID</th>
<th>Barrier Category</th>
<th>Trail ID</th>
<th>Distance (m)</th>
<th>Road ID</th>
<th>Distance (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fires Creek</td>
<td>U-FC4</td>
<td>MODERATE</td>
<td>TR389</td>
<td>66.9</td>
<td>340</td>
<td>56.1</td>
</tr>
<tr>
<td></td>
<td>U-FC1</td>
<td>MODERATE</td>
<td>TR389</td>
<td>68.9</td>
<td>340</td>
<td>11.3</td>
</tr>
<tr>
<td></td>
<td>U-L FC6</td>
<td>MODERATE</td>
<td>TR78</td>
<td>435.4</td>
<td>340</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>U-L FC5</td>
<td>MODERATE</td>
<td>TR78</td>
<td>259.2</td>
<td>340</td>
<td>3.8</td>
</tr>
<tr>
<td></td>
<td>FC2</td>
<td>MODERATE</td>
<td>TR78</td>
<td>556.4</td>
<td>340</td>
<td>4.8</td>
</tr>
<tr>
<td></td>
<td>FC1</td>
<td>MODERATE</td>
<td>TR386</td>
<td>706.9</td>
<td>340</td>
<td>6.1</td>
</tr>
<tr>
<td></td>
<td>LBC2</td>
<td>MAJOR</td>
<td>TR80</td>
<td>294.0</td>
<td>340C</td>
<td>20.9</td>
</tr>
<tr>
<td></td>
<td>LBC1</td>
<td>MODERATE</td>
<td>TR80</td>
<td>370.3</td>
<td>340C</td>
<td>12.1</td>
</tr>
<tr>
<td></td>
<td>L-LBC10</td>
<td>MAJOR</td>
<td>TR78</td>
<td>1256.4</td>
<td>340C</td>
<td>5.8</td>
</tr>
<tr>
<td></td>
<td>L-LBC9</td>
<td>MODERATE</td>
<td>TR78</td>
<td>951.4</td>
<td>340C</td>
<td>1.7</td>
</tr>
<tr>
<td></td>
<td>L-LBC7</td>
<td>MODERATE</td>
<td>TR78</td>
<td>805.6</td>
<td>340C</td>
<td>5.0</td>
</tr>
<tr>
<td></td>
<td>L-LBC6</td>
<td>MODERATE</td>
<td>TR78</td>
<td>725.8</td>
<td>340C</td>
<td>9.2</td>
</tr>
<tr>
<td></td>
<td>L-LBC4</td>
<td>MAJOR</td>
<td>TR78</td>
<td>897.1</td>
<td>340C</td>
<td>1.7</td>
</tr>
<tr>
<td></td>
<td>L-LBC2</td>
<td>MODERATE</td>
<td>TR78</td>
<td>292.1</td>
<td>340C</td>
<td>0.3</td>
</tr>
<tr>
<td>Long Branch</td>
<td>RC1</td>
<td>MODERATE</td>
<td>TR387</td>
<td>9.6</td>
<td>340A2</td>
<td>19.0</td>
</tr>
<tr>
<td></td>
<td>RC5</td>
<td>MODERATE</td>
<td>TR387</td>
<td>188.2</td>
<td>340A2</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>L-RC2</td>
<td>MODERATE</td>
<td>TR388</td>
<td>692.0</td>
<td>340A</td>
<td>1.4</td>
</tr>
<tr>
<td></td>
<td>L-RC1</td>
<td>MODERATE</td>
<td>TR74</td>
<td>682.0</td>
<td>340A</td>
<td>2.2</td>
</tr>
<tr>
<td>Rockhouse Creek</td>
<td>LC1</td>
<td>MODERATE</td>
<td>TR388</td>
<td>98.6</td>
<td>340A1</td>
<td>0.4</td>
</tr>
<tr>
<td></td>
<td>LC2</td>
<td>MAJOR</td>
<td>TR388</td>
<td>30.6</td>
<td>340A1</td>
<td>1.9</td>
</tr>
<tr>
<td></td>
<td>LC3</td>
<td>MODERATE</td>
<td>TR388</td>
<td>89.4</td>
<td>340A1</td>
<td>9.2</td>
</tr>
<tr>
<td>Laurel Creek</td>
<td>U-LFC1</td>
<td>MODERATE</td>
<td>TR386</td>
<td>1.8</td>
<td>340B</td>
<td>0.2</td>
</tr>
<tr>
<td></td>
<td>LFC1</td>
<td>MODERATE</td>
<td>TR386</td>
<td>0.4</td>
<td>340B</td>
<td>0.0</td>
</tr>
<tr>
<td>Little Fires Creek</td>
<td>U-LFC1</td>
<td>MODERATE</td>
<td>TR386</td>
<td>1.8</td>
<td>340B</td>
<td>0.2</td>
</tr>
<tr>
<td></td>
<td>LFC1</td>
<td>MODERATE</td>
<td>TR386</td>
<td>0.4</td>
<td>340B</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>BSB1</td>
<td>MAJOR</td>
<td>TR78</td>
<td>107.8</td>
<td>340</td>
<td>22.6</td>
</tr>
<tr>
<td></td>
<td>Collett Camp Branch</td>
<td>340C+CCB</td>
<td>TR72</td>
<td>1159.3</td>
<td>340C</td>
<td>0.9</td>
</tr>
<tr>
<td></td>
<td>Tom Cove Branch</td>
<td>TC1</td>
<td>TR74</td>
<td>83.5</td>
<td>340</td>
<td>18.6</td>
</tr>
<tr>
<td></td>
<td>Leatherwood Branch</td>
<td>LWD FALL</td>
<td>TR73</td>
<td>9.1</td>
<td>340E</td>
<td>11.7</td>
</tr>
</tbody>
</table>
Figure 6. The map shows the distribution and genetic identity of brook trout populations in the Hiwassee River Basin.
III. Development of a Model to Predict Migration Barriers

As previously described, migration barriers within a stream network are highly variable structures. They occur as debris jams, massive waterfalls, boulder clusters and as a variety of other physical structures. To identify environmental variables that indicate barriers, a classification tree model is developed using barrier data from the Fires Creek Watershed.

Two final models are derived. The first depicts the response variable as 0 and 1, or non-barrier (n=100) and barrier (n=26). This model is referred to as MR1 (Figure 7A). The second separates the response according to barrier types ‘Moderate’ and ‘Major’. The response categories are then 0, 1, and 2, which respectively correspond to non-barrier (n=100), ‘Moderate’ barrier (n=19), and ‘Major’ barrier (n=7). This model is labeled MR2 (Figure 7B).

Model accuracy, sensitivity (true positive rate; TPR), and specificity (true negative rate; TNR) is calculated for each fitted model from a confusion matrix (Figures 8A and 8B). Results show that overall accuracy is similar for both but that model sensitivity declines when responses are separated according to barrier type. This indicates that decreasing the presence/absence ratio in the sample data by sub-dividing the barrier response group diminishes the classification success of the model.

Variable pathways in MR1 that lead to a barrier response are mapped in the Fires Creek Watershed and predictions are compared with known barrier locations (Figure 9 and 10). In contrast to the classification success derived from the fitted model, which consider 26 presence points and a sub-sample (n=100) of the 11,973 absence points, mapped results are assessed over the entire stream network (presence = 26 and absence = 11,973). However, approximately 2,600 absence points are unconfirmed. These correspond to habitat in reaches above the ‘Major’
barriers on Leatherwood Falls, Tom Cove Branch, and Baldspring Branch in addition to the entire length of Tatham Cabin and Hickory Cove Creeks.

Values for accuracy, sensitivity (TPR), specificity (TNR), a false positive rate (FPR), and a false negative rate (FNR) are calculated for mapped predictions. Results suggest that mapped accuracy and sensitivity are identical to measures of classification success derived from the CART model. The true negative rate, however; declines from 90% to 83% indicating less specificity when model outputs are applied over the full extent of the stream network. If predicted barrier locations are mapped using the model that distinguishes between ‘Major’ and ‘Minor’ barriers, mapping accuracy severely declines in comparison to values indicated in CART model outputs (data not shown).

False negative and false positive locations are distinguished to examine the spatial distribution of such errors and to identify omitted barriers. False negatives (FNR = 23%) occur at four moderate (U-FC1, U-LFC6, L-LBC7, L-LBC9) and two major barriers (L-LBC10 and TC1). The false positive rate is relatively low, i.e. 17%, but the extent of false positive locations suggests that the model is over-fitted and tree-pruning is necessary to improve sensitivity.

CART model results from MR1 are applied to the Cold Spring Creek Watershed using watershed-specific variables to predict barriers. The validity of applying CART outputs derived from data in the Fires Creek Watershed is upheld by confirming that the mean and range of predictor variables is similar (Table 7). The mapped results demonstrate that numerous locations are predicted to contain barriers (data not shown). Model predictions are validated by data collected in Cherry Creek and the upper section of Cold Spring Creek. Overall, classification success declines in this watershed; the model correctly predicts only 4 out of 11 barriers found in Cherry Creek and only 1 out of 8 in upper-Cold Spring Creek (Figures 11 and 12).
Figure 7. The hierarchical tree diagrams for Model Run 1 and 2 show the split and sequence of environmental variables that lead to a particular response (1 = barrier, 0 = non-barrier). At each branch, the response is classified by the split of a particular environmental variable or by the interaction of preceding variables. Variables include: 

- focdif5 = 5x5 focal max elevation – focal min elevation;
- focdif3 = 3x3 focal max elevation – focal minimum elevation;
- ccmn = 3x3 focal mean of percent canopy cover;
- slp = slope;
- strslp = stream gradient;
- mn5ed = focal 5x5 mean – actual elevation;
- mn3ed = focal 3x3 mean – actual elevation.

(see Table 2).
A. Model Run 1  

<table>
<thead>
<tr>
<th>Actual Values</th>
<th>0</th>
<th>1</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predicted Values</td>
<td>0</td>
<td>90</td>
<td>6</td>
</tr>
<tr>
<td>1</td>
<td>10</td>
<td>20</td>
<td>30</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>100</td>
<td>26</td>
<td>126</td>
</tr>
</tbody>
</table>

Accuracy: 87%
TPR: 77%
TNR: 90%

B. Model Run 2  

<table>
<thead>
<tr>
<th>Actual Values</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predicted Values</td>
<td>0</td>
<td>92</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>6</td>
<td>14</td>
<td>2</td>
<td>22</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>0</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>100</td>
<td>19</td>
<td>7</td>
<td>126</td>
</tr>
</tbody>
</table>

Accuracy: 86%
TPR: 65%
TNR: 92%

**TPR (1) and TPR (2)** indicate true positive rates specific to the respective barrier response of 1 and 2.

**Figure 8.** Confusion matrices show predicted responses derived from the fitted models versus actual responses; (a) Model Run 1 and (b) Model Run 2. The distribution of misclassification errors are used to calculate accuracy, sensitivity or TPR, and specificity or TNR. (a) Model Run 1 depicts the response variable as 0 and 1; respectively, non-barrier \( n=100 \) and barrier \( n=26 \). (b) Model Run 2 includes the response variable separated according to barrier type; 0 = non-barrier \( n=100 \), 1 = ‘Moderate’ barriers \( n=19 \), and 2 = ‘Major’ \( n=7 \) barriers. TPR (1) and TPR (2) indicate true positive rates specific to the respective barrier response of 1 and 2.

**Model Run 1 - Mapped Results**  

<table>
<thead>
<tr>
<th>Known Barriers (pixels)</th>
<th>0</th>
<th>1</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mapped Barriers (pixels)</td>
<td>0</td>
<td>9887</td>
<td>6</td>
</tr>
<tr>
<td>1</td>
<td>2086</td>
<td>20</td>
<td>2106</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>11973</td>
<td>26</td>
<td>11999</td>
</tr>
</tbody>
</table>

Accuracy: 83%
TPR: 77%
FNR: 23%
TNR: 83%
FPR: 17%

**Figure 9.** Classification success measures are calculated from the confusion matrix, which displays the number of predicted versus known barriers in the Fires Creek Watershed. Results are derived from mapping environmental variables and interaction pathways that lead to a barrier response (i.e. response = 1) in Model Run 1 (FPR = false positive rate, FNR = false negative rate).
Figure 10. The map illustrates the predicted locations of barriers within the watershed and demonstrates the number of true and false predictions.
Table 7. Predictor variable values for the Fires Creek and the Cold Spring Creek Watersheds are compared to support the validity of applying a CART model derived from Fires Creek to a different watershed.

<table>
<thead>
<tr>
<th>Watershed</th>
<th>Fires Creek</th>
<th>Cold Spring Creek</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variables</td>
<td>Mean Value (Range)</td>
<td>Mean Value (Range)</td>
</tr>
<tr>
<td>Elevation*</td>
<td>3001 (1614 - 5267)</td>
<td>3278 (1850 - 4408)</td>
</tr>
<tr>
<td>Slope</td>
<td>25 (0 - 75)</td>
<td>23 (0 - 59.8)</td>
</tr>
<tr>
<td>Mn5ed (Mean elev - elev)</td>
<td>0 (-44.2 - 68.4)</td>
<td>0 (-19.8 - 37)</td>
</tr>
<tr>
<td>Mn3ed (Mean elev - elev)</td>
<td>0 (-30.6 - 43.3)</td>
<td>0 (-16.3 - 29.4)</td>
</tr>
<tr>
<td>FocDif5 (Max elev - min elev)</td>
<td>49 (0 - 262)</td>
<td>44 (0 - 138)</td>
</tr>
<tr>
<td>FocDif3 (Max elev - min elev)</td>
<td>25 (0 - 187)</td>
<td>22 (0 - 93)</td>
</tr>
<tr>
<td>Stream gradient</td>
<td>0.103 (0.008 - 0.44)</td>
<td>0.130 (0.002 - 0.25)</td>
</tr>
<tr>
<td>Canopy Cover (focal mean)</td>
<td>91 (0 - 100)</td>
<td>89 (0 - 100)</td>
</tr>
</tbody>
</table>

*Elevation is not included in the model but is used to derive slope and the local relief variables.

Model Run 1: Mapped Results (Cold Spring Creek Watershed)

<table>
<thead>
<tr>
<th>Cherry Creek</th>
<th>Upper Cold Spring Creek</th>
</tr>
</thead>
<tbody>
<tr>
<td>Known Barriers (pixels)</td>
<td>Mapped Barriers (pixels)</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>461</td>
</tr>
<tr>
<td>1</td>
<td>106</td>
</tr>
<tr>
<td>Total</td>
<td>567</td>
</tr>
</tbody>
</table>

Accuracy 80% 79%

<table>
<thead>
<tr>
<th>TPR</th>
<th>TNR</th>
<th>FNR</th>
<th>FPR</th>
</tr>
</thead>
<tbody>
<tr>
<td>73%</td>
<td>81%</td>
<td>27%</td>
<td>19%</td>
</tr>
<tr>
<td>87%</td>
<td>20%</td>
<td>13%</td>
<td>80%</td>
</tr>
</tbody>
</table>

Figure 11. Classification success measures are calculated from confusion matrices that show for each stream, barriers predicted by mapping pathways in Model Run 1 that lead to responses = 1 and known locations derived from field work.
Figure 12. The map demonstrates the location of predicted barriers in Cherry Creek and upper-Cold Spring Creek.
**DISCUSSION**

*Implications for Brook Trout Restoration in the Fires Creek Watershed*

The identification of ‘Major’ and ‘Moderate’ barriers serves as a framework for further exploring the viability of the Fires Creek Watershed as a site for brook trout restoration. Almost all of the surveyed reaches, except for 10 kilometers on the main Fires Creek channel, contain completely impassable, ‘Major’, or variably passable, ‘Moderate’, structures. Approximately 20 kilometers of potential habitat are located above ‘Major’ barriers; however, not all of these upstream reaches have been explored and some are interrupted by additional ‘Major’ barriers (e.g. Long Branch).

Unlike habitat located above ‘Major’ barriers, reaches above ‘Moderate’ structures are not guaranteed full protection from non-native salmonid migration. For example, rainbow trout were observed upstream of several ‘Moderate’ barriers on the main Fires Creek channel. However, opportunities exist to reinforce the integrity of these barriers with man-made structures such as rock gabions, to prevent upstream movement of non-native trout (Thompson and Rahel 1998). Such measures would increase the amount of definitively protected habitat and offer greater flexibility in the design of a restoration strategy.

A minimum patch size for brook trout habitat has yet to be determined but research suggests that seasonal variations and life cycle stages influence habitat requirements (Whitworth and Strange 1983; Petty et al. 2005). Habitat modeling of bull trout demonstrates that large scale spatial processes affect population persistence (Reiman and McIntrye 1995). In addition, habitat restricted to headwater areas is often susceptible to extreme events such as flooding or drought (Larson and Moore 1985). Collectively, this evidence suggests that survival of reintroduced brook trout populations is improved by providing diverse habitat options.
Both the presence of rainbow trout and water quality parameters indicated by temperature and pH, demonstrate that the Fires Creek Watershed offers adequate habitat for brook trout. However, the distribution of rainbow trout along the main Fires Creek channel suggests that successful restoration activities will entail extensive elimination or relocation of extant rainbow trout. In addition, future environmental changes affect the long-term suitability of available habitat within the watershed.

Stream temperature increases predicted by global climate change will affect efforts to restore brook trout populations to cold, headwater systems in the Southern Appalachians. Models show that stream warming decreases the amount of available brook trout habitat by increasing the minimum elevation at which it is found (Meisner 1990; Flebbe 2006).

Specifically, changes in stream temperature modify the solubility of gases, rates of decomposition, and the metabolism, behavior, growth, and survival of aquatic organisms (Mitchell 1999; Johnson and Jones 2000). Warmer waters decrease resistance to predation and disease, inhibit feeding and reproduction, and alter intra-specific competition. Stream temperatures become lethal to brook trout at or above 24°C (Swift and Messer 1971; Raleigh 1982). Further habitat modifications caused by increased stream temperatures include altered hydrologic flows and changes in riparian vegetation types (Flebbe 1997).

While brook trout exhibit tolerance to a wide range of pH values, optimal growth occurs between 6.8 and 8 (Raleigh 1982). Episodic stream acidification decreases the viability of embryos and survival of the earliest life stages (Fiss and Carline 1993; Gagen et al. 1993). At high elevations, streams are susceptible to acidity deposited by anthropogenic pollution and geologic weathering (SAMAB 1996). Both sources potentially limit the extent of suitable habitat. Cursory surveys of stream pH suggest that acidity is not an issue for the Fires Creek
Watershed; however, the potential for future contributions from anthropogenic sources is unknown.

Interestingly, acidification and warming impacts move suitable habitat in opposite directions. Appropriate habitat under increased acidification occurs below 1,067m while in response to global warming, it occurs above 915m (Habera and Moore 2005). These considerations converge onto a narrow range of suitable habitat located between 915 and 1,067m. Currently, 26% or 15 kilometers of the maximum available habitat in the Fires Creek Watershed occurs within these elevations. For the purposes of targeting specific sites within the watershed, it is important to note that 52.2% or 3.2 miles of the longest reach on Fires Creek and 47.4% or 2.4 miles of the longest reach on Long Branch falls within this range.

Efforts to reintroduce brook trout to the Fires Creek Watershed are facilitated by the proximate location of suitable donor populations. Restoration of SABT requires that donors demonstrate the unique genetic lineage associated only with brook trout in the Southern Appalachians (McCracken et al. 1993; Hayes et al. 1996; Habera and Moore 2005). Also, SABT located within the immediate geographic region are preferred to maintain genetic variability between geographically separated populations; selecting nearby populations also minimizes the impact of transport. Although brook trout are not abundant in the region, genetically referenced fish distribution data show a suitable donor population in a neighboring watershed, southeast of Fires Creek.

The location of roads and trails with respect to potential reintroduction sites is a seemingly minor but very important detail. As calculated by the nearest distance to each barrier, roads are closer than trails. However, road conditions in the watershed are variable and in some cases, impassable by car. An investigation of the proposed travel path prior to the actual
reintroduction is necessary to ensure that humans and fish travel the least cost path to selected sites. Also, it is important to consider the proximity of roads to streams as a potential source of sedimentation and habitat impairment.

**Implications for Model Development and Applicability**

A CART model is developed from barrier data in the Fires Creek Watershed to predict the location of migration barriers. When CART outputs are mapped into the watershed from which it is derived, results show that the model misses six known barriers and generally over-predicts locations throughout the watershed. Given that the purpose of the model is to improve the efficiency at which barriers are located in candidate watersheds, the mapped false positive rate of 17% is prohibitive and suggests that model improvements are necessary to increase the usefulness of the CART model.

**CART Results in R**

It is important to note that running the CART algorithm with different subsets of randomly sampled zeros (\( n = 100 \)) alters the output. The raw data contain a dominant number of absences (\( n = 11,973 \)), and thus each time zeros are randomly sub-sampled, the values of associated predictors are potentially variable and the algorithm produces a different tree output. In general, a good model is robust to variations associated with randomly sub-sampled zeros but CART methods always produce models that are over-fitted to the input data.

In this case, when models are created from different subsets of zeros (\( n = 100 \)), variations are present at the level of variable selection and/or the split at which a particular variable is included. Therefore, to accurately compare models with and without certain variables or to assess classification measures between original fitted trees and pruned trees, model variations must be derived from the same sub-sampling of data. Alternatively, a more robust model may be
developed by first determining the optimal sub-sample number through evaluation of models constructed with successively larger sub-samples of zeros.

However, regardless of the zero subset \((n=100)\) included in the model, the following variables appear in every CART model derived: 5x5 focal maximum elevation minus focal minimum elevation \((foc5dif)\), slope \((slp)\), 3x3 focal mean of canopy cover \((ccmn)\), and 5x5 focal elevation mean minus actual elevation \((mn5ed)\). Slope and local relief variables (indicated by focal elevation manipulations) capture aspects of terrain variability that contribute to barrier occurrences within a stream network. The focal mean of canopy cover depicts a measure of openness surrounding a structure and potentially indicates clearings associated with large ‘Major’ barriers. Overall, model results suggest that these variables are major determinants of barriers.

A potentially unavoidable problem arises from inaccuracies in the GPS waypoints, which subsequently create error in the model response variable. Points are taken at each in-stream barrier location or at the nearest location for which satellite reception is available. These points are subsequently translated into locations within the NHD Plus derived stream network \((1:100,000\) scale) by designating the raster cells closest to each waypoint as a barrier. The stream raster, coded to indicate barrier locations, is then used to sample values from predictor variables. Through this process, positional inaccuracies derived from the GPS are transferred to the medium resolution stream raster, and thus error is propagated into the designated position of barriers and corresponding predictor values. These errors may be mitigated by representing barrier locations in the response raster over a larger spatial extent.
Mapped Results

Mapping model results in a GIS provides an opportunity to improve the overall model by examining the spatial distribution of predicted barrier locations. For example, identifying specific characteristics associated with false negatives in the Fires Creek Watershed reveals that four of the six missed barriers are long sloping features. This suggests that it may be beneficial to depict certain barriers over a larger spatial extent instead of representing each barrier in the response by only a single pixel.

In the Cold Spring Creek Watershed, model success declines in comparison to its performance in Fires Creek. Yet, important information regarding model performance is gained by validating barrier predictions in Cherry Creek and upper Cold Spring Creek. Results indicate that overall model success in Cherry Creek is higher than in upper Cold Spring Creek. These findings correspond to striking differences in stream morphology and riparian structure; Cherry Creek is a high gradient, wider stream that contains several distinct barrier features while in contrast, upper Cold Spring Creek is relatively narrow with shallow flows, low gradient areas, and numerous indistinct debris filled blockages. Thus, the model performs better in larger streams that contain well-defined barrier features; this is a reasonable expectation and conducive to the goal of identifying barriers within suitable habitat.

The model in its current specification (MR1) is over-fitted to the data from which it is constructed, i.e. the barrier data from the Fires Creek Watershed. This diminishes its robustness in new watersheds, as demonstrated by the results derived from the Cold Spring Creek Watershed. However, attempts to refine the model specification by cross-validation and tree-pruning procedures (i.e. decrease the number of branches to an optimal number indicated by
cross-validation results) do not show improved classification success in either the Fires Creek or Cold Spring Creek Watersheds (data not shown).

Overall, model success is limited by the lack of highly resolved, widely available, and user-friendly data. Predictor variables derived from high resolution datasets are more likely to capture the finely detailed, non-homogenous features that distinguish barriers in a stream network. Factors such as stream slope, which in conjunction with other landscape features are likely to be good indicators of barriers, are currently available only at the level of an individual catchment (i.e. per unit of uninterrupted reach). Geology is another potentially useful predictor variable that is only available at large spatial scales. The composition of bedrock underlying streams and riparian areas influences stream morphology and potentially indicates areas that are susceptible to water erosion and/or boulder deposition. Respectively, these areas might contribute to waterfalls and/or chutes, and rock blockages. Finally, data that accurately depicts stream width and bank slope are not readily available, yet together; these variables possibly reflect areas of high flow or hydropower associated with sharp drops as in waterfalls and/or cascades.
CONCLUSIONS

Ultimately, the most effective restoration strategy combines the availability of protected habitat delineated by barrier identification with information extracted from the comprehensive assessment of the Fires Creek Watershed. Based on this information, habitat above two ‘Moderate’ barriers along the main Fires Creek channel represents an optimal target for brook trout reintroduction. As previously discussed, removal of rainbow trout and barrier reinforcement is necessary. However, these reaches offer access to diverse habitat, which helps to ensure the survival of reintroduced populations.

Alternatively, a potentially less costly and labor intensive approach is to target areas directly above the first ‘Major’ barrier on Long Branch. This would eliminate the need to reinforce barriers along Fires Creek. Reducing the total habitat available to reintroduced populations also facilitates monitoring and thus, the approach might serve as a pilot study to assess the viability of brook trout restored to the Fires Creek Watershed.

Finally, attempts to model the location of migration barriers within the stream network suggest that more highly resolved data are necessary to improve the usefulness of the model. Currently, the extent of mapped false positives impedes the model’s ability to improve the efficiency at which trout migration barriers are identified in candidate watersheds. However, the model provides a framework for further development of a tool that streamlines the barrier identification process.
ACKNOWLEDGEMENTS

I would first like to thank Hugh Irwin, the Conservation Planner at SAFC, for providing me with an invaluable learning experience and an opportunity to engage in exciting conservation work. Also, I am grateful for the support and assistance provided by Mark Shelley, Nicole Martinez, Kim Porter, and Josh Kelley at SAFC. Additionally, the following people contributed to my summer work in Asheville, NC: Sheryl Bryan, Fisheries Biologist for North Carolina’s National Forests, Jason Farmer, Fisheries Biologist for Nantahala National Forest, and Dr. Bill McLarney.

This project would not have been possible without the enduring support of faculty, staff, and students at the Nicholas School. In particular, I am indebted to my patient and wise advisor, Dr. Dean Urban, for his constant advice and guidance. I would also like to thank Ben Best, who was instrumental in developing the modeling strategy and Pete Harrell for all of his friendly GIS assistance.
REFERENCES


Meisner J.D. 1990. Effect of climatic warming on the southern margins of the native range of brook trout, Salvelinus fontinalis. Canadian Journal of Fisheries and Aquatic Sciences 47: 1065-1070


NC Department of Environment and Natural Resources – Division of Parks and Recreation: Natural Heritage Program. Raleigh, NC. http://www.ncnhp.org


Appendix A: GIS Data Sources

This appendix provides more detailed information regarding the GIS data and sources used in the watershed assessment and also model development.

Table A-1. GIS Data Sources

<table>
<thead>
<tr>
<th>Feature</th>
<th>Dataset</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fish Distribution</td>
<td>Fish survey data for Hiwassee River Basin</td>
<td>Jason Farmer and Sheryl Bryan with USFS-National Forests in NC</td>
</tr>
<tr>
<td>Site Accessibility</td>
<td>Forest Service road and trail shapefiles for the Fires Creek Watershed</td>
<td>Sheryl Bryan with USFS-National Forests in NC</td>
</tr>
<tr>
<td>Genetics</td>
<td>Genetic results for Hiwassee River Basin provided in spreadsheet format but subsequently linked to spatial fish distribution data.</td>
<td>Doug Besler with the NCWRC</td>
</tr>
<tr>
<td>Elevation</td>
<td>LiDAR derived digital elevation model for Clay and Haywood Counties, NC (Raster cell size = 20 feet)</td>
<td>North Carolina Department of Transportation: <a href="http://www.ncdot.org/it/gis/DataDistribution/ContourElevation">http://www.ncdot.org/it/gis/DataDistribution/ContourElevation</a></td>
</tr>
<tr>
<td>Soil</td>
<td>Soil polygons derived at 1:12,000 Soil percent clay and erodibility factors obtained from the provided Microsoft Access database</td>
<td>USDA - Natural Resource Conservation Service: <a href="http://soildatamart.nrcs.usda.gov">http://soildatamart.nrcs.usda.gov</a></td>
</tr>
<tr>
<td>Canopy Cover</td>
<td>2001 National land cover data (Raster cell size = 30m)</td>
<td>USGS: <a href="http://gisdata.usgs.net/website/MRLC">http://gisdata.usgs.net/website/MRLC</a></td>
</tr>
</tbody>
</table>
APPENDIX B: ADDITIONAL STREAM HABITAT AND BARRIER NOTES

Leatherwood Creek:
- Fires Creek tributary.
- Potential trout habitat above the fall, which is located very close to the confluence with Fires Creek, was not assessed.

Rockhouse Creek:
- Fires Creek tributary.
- Road 340A follows the creek until the confluence with Laurel Creek which is located beneath a culvert at the point where the road is blocked by a gate.
- This culvert is likely impassable but this needs to be reconfirmed.
- A trail exists most of the distance up to barrier RC1.
- At barrier RC1, average stream width is 8 feet. Sighting upstream from this location, the width appears to become increasingly narrow.
- Stream was explored up to waypoint RC1.
- Impression is that suitable habitat exists along this stream between marked barriers and at least up to waypoint RC1.

Laurel Creek:
- Rockhouse Creek tributary.
- Road 340A, above the confluence with Rockhouse Creek and past the road gate, follows very close to the stream and many sediment fences are failing.
- A gulley exists along the side of the road opposite the stream, but it does not extend the entire length.
- Above barrier LC4, the stream begins to flatten and widen, creating classic jump/rest habitat.
- Stream was explored up to waypoint LC6.
- Impression is that suitable habitat exists between the barriers on this stream and up until at least waypoint LC6.

Tom Cove:
- Fires Creek tributary.
- Potential trout habitat above the cascade, which is located at the confluence with Fires Creek, was not assessed.

Little Fires Creek:
- Fires Creek tributary.
- Road 340B parallels this stream.
- The furthest distance traveled upstream was 1.8 road miles.
- The road crosses the stream at 0.7 miles. LFC1 exists below this point and U-LFC1 exists above.
- The reach above the road crossing, waypoint 340B-X, is very narrow and totally covered in rhododendron.
• Impression is that the stream is probably less productive or fertile here and that trout habitat would be marginal.

Baldspring Branch:
• Fires Creek tributary.
• This creek is narrow and rocky beginning at the confluence with Fires Creek and up to a huge waterfall that extends up the slope for an indeterminable distance.
• Impression is that stream is not a candidate for restoration.

Long Branch Creek:
• Largest Fires Creek tributary.
• Road 340C parallels this stream up to waypoint, 340C end.
• The full extent of the stream was surveyed and access to the uppermost sections is available via the main Fires Creek road, 340.
• The lower reach of this stream is characterized by yellow birch and rhododendron vegetation and small black snails cover the rocks.
• Impression is that the stream has many striking features that would serve as potential barriers and the habitat quality seems very high.

Collett Camp Branch:
• Long Branch Creek tributary located between roads 427 and 6274.
• The confluence with Long Branch occurs below the point (Waypoint 340C+CCB) where road 340C crosses the stream.
• Two impassable culverts empty into the downstream side.
• Above the road crossing, stream is mostly bedrock and narrow and quickly becomes a series of cascades with increasing slope. Some pools are interspersed with a maximum width of 4 feet.
• Overall impression is that above the road crossing, habitat may exist but it is not high quality.

Coldspring Branch:
• Long Branch Creek tributary located near the turnoff onto road 340C and before road 427.
• The confluence with Long Branch occurs below the point (waypoint 340C+CB) where road crosses the stream.
• A single culvert exists that is failing on the upstream side of the road. Cinder blocks are partially blocking the opening. The collapse of the supporting walls seems imminent.
• The stream above the road crossing is mostly a series of long sloping bedrock cascades with very small or no pools.
• It is further characterized by predominantly red rock bedrock with interspersed black bands.
• Much of the exposed or partially submerged surface of the bedrock is covered in hornwort.
Fires Creek

- Road 340 parallels most of the stream.
- The entire reach between the Fires Creek picnic area and the barrier UFC4 was surveyed.
- The reach above the bridge on 340B to the confluence with Long Branch is characterized by bedrock streambeds and many long, mildly sloping bedrock cascades.
- Overall impression is that suitable habitat exists along most of the stream.
Figure C-1. Sampling values from predictor variables to generate CART input data.
Figure C-2. CART tree branches leading to barrier responses are mapped as conditional statements to spatially demonstrate predictions and compare with known locations in Fires Creek and Cold Spring Creek Watersheds.
**Python Scripts:**

**Figure C-1**

```python
# Sample_model.py
# Created on: Sun Apr 13 2008 10:13:50 PM
# (generated by ArcGIS/ModelBuilder)

# Import system modules
import sys, string, os, arcgisscripting

# Create the Geoprocessor object
gp = arcgisscripting.create()

# Check out any necessary licenses
gp.CheckOutExtension("spatial")

# Load required toolboxes...
gp.AddToolbox("C:/Program Files/ArcGIS/ArcToolbox/Toolboxes/Spatial Analyst Tools.tbx")
gp.AddToolbox("C:/Program Files/ArcGIS/ArcToolbox/Toolboxes/Conversion Tools.tbx")

# Set the Geoprocessing environment...
gp.XYResolution = 

# Local variables...
EX_elev_FC = "z:\SAFC_INFO\\LIDAR from NCDOT\Lidar2007_Clay\LIDAR DEMs.mdb\EX_elev_FC"
Slope2_FC = "z:\SAFC_INFO\\LIDAR from NCDOT\Lidar2007_Clay\LIDAR DEMs.mdb\Slope2_FC"
FocMn5_elevdif = "z:\SAFC_INFO\\LIDAR from NCDOT\Lidar2007_Clay\LIDAR DEMs.mdb\FocMn5_elevdif"
FocMn3_elevdif = "z:\SAFC_INFO\\LIDAR from NCDOT\Lidar2007_Clay\LIDAR DEMs.mdb\FocMn3_elevdif"
FocDif5_FC = "z:\SAFC_INFO\\LIDAR from NCDOT\Lidar2007_Clay\LIDAR DEMs.mdb\FocDif5_FC"
FocDif3_FC = "z:\SAFC_INFO\\LIDAR from NCDOT\Lidar2007_Clay\LIDAR DEMs.mdb\FocDif3_FC"
nhdfl_FC_strmslp = "z:\SAFC_INFO\\LIDAR from NCDOT\Lidar2007_Clay\LIDAR DEMs.mdb\nhdfl_FC_strmslp"
Catch_NLCD41_2 = "z:\SAFC_INFO\\LIDAR from NCDOT\Lidar2007_Clay\LIDAR DEMs.mdb\Catch_NLCD41_2"
Catch_NLCD42_2 = "z:\SAFC_INFO\\LIDAR from NCDOT\Lidar2007_Clay\LIDAR DEMs.mdb\Catch_NLCD42_2"
Catch_NLCD43_2 = "z:\SAFC_INFO\\LIDAR from NCDOT\Lidar2007_Clay\LIDAR DEMs.mdb\Catch_NLCD43_2"
can_FC_focmn = "z:\SAFC_INFO\\LIDAR from NCDOT\Lidar2007_Clay\LIDAR DEMs.mdb\can_FC_focmn"
can_FC_focmax = "z:\SAFC_INFO\\LIDAR from NCDOT\Lidar2007_Clay\LIDAR DEMs.mdb\can_FC_focmax"
```

53
nhdfl_pts_26 = "nhdfl_pts_26"
Sample_26 = "z:\SAFC_INFO\LIDAR from NCDOT\Lidar2007_Clay\LIDAR DEMs.mdb\Sample_26"
Sample_26_1 = "z:\SAFC_INFO\LIDAR from NCDOT\Lidar2007_Clay\LIDAR DEMs.mdb\Sample_26_1"
soil_K = "soil_K"
soil_clay = "soil_clay"
Sample_26_2_1 = "z:\SAFC_INFO\LIDAR from NCDOT\Lidar2007_Clay\LIDAR DEMs.mdb\Sample_26_2_1"
Samp_26_2_1 = "z:\SAFC_INFO\LIDAR from NCDOT\Lidar2007_Clay\LIDAR DEMs.mdb\Samp_26_2_1"

# Process: Point to Raster (2)...
gp.PointToRaster_conversion(nhdfl_pts_26, "Blockage", Sample_26, "MOST_FREQUENT", "NONE", "20")

# Process: Sample (2)...
gp.Sample_sa("'z:\SAFC_INFO\LIDAR from NCDOT\Lidar2007_Clay\LIDAR DEMs.mdb\EX_elev_FC';'z:\SAFC_INFO\LIDAR from NCDOT\Lidar2007_Clay\LIDAR DEMs.mdb\Slope2_FC';'z:\SAFC_INFO\LIDAR from NCDOT\Lidar2007_Clay\LIDAR DEMs.mdb\FocMn5_elevdif';'z:\SAFC_INFO\LIDAR from NCDOT\Lidar2007_Clay\LIDAR DEMs.mdb\FocMn3_elevdif';'z:\SAFC_INFO\LIDAR from NCDOT\Lidar2007_Clay\LIDAR DEMs.mdb\FocDif5_FC';'z:\SAFC_INFO\LIDAR from NCDOT\Lidar2007_Clay\LIDAR DEMs.mdb\FocDif3_FC';'z:\SAFC_INFO\LIDAR from NCDOT\Lidar2007_Clay\LIDAR DEMs.mdb\nhdfl_FC_strmslp';'z:\SAFC_INFO\LIDAR from NCDOT\Lidar2007_Clay\LIDAR DEMs.mdb\Catch_NLCD41_2';'z:\SAFC_INFO\LIDAR from NCDOT\Lidar2007_Clay\LIDAR DEMs.mdb\Catch_NLCD42_2';'z:\SAFC_INFO\LIDAR from NCDOT\Lidar2007_Clay\LIDAR DEMs.mdb\Catch_NLCD43_2';'z:\SAFC_INFO\LIDAR from NCDOT\Lidar2007_Clay\LIDAR DEMs.mdb\can_FC_focmn';'z:\SAFC_INFO\LIDAR from NCDOT\Lidar2007_Clay\LIDAR DEMs.mdb\can_FC_focmax'", Sample_26, Sample_26_1, "NEAREST")

# Process: Point to Raster (3)...
gp.PointToRaster_conversion(nhdfl_pts_26_2, "Blockage", Sample_26_2_1, "MOST_FREQUENT", "NONE", "20")

# Process: Sample (5)...
gp.Sample_sa("'z:\SAFC_INFO\LIDAR from NCDOT\Lidar2007_Clay\LIDAR DEMs.mdb\EX_elev_FC';'z:\SAFC_INFO\LIDAR from NCDOT\Lidar2007_Clay\LIDAR DEMs.mdb\Slope2_FC';'z:\SAFC_INFO\LIDAR from NCDOT\Lidar2007_Clay\LIDAR DEMs.mdb\FocMn5_elevdif';'z:\SAFC_INFO\LIDAR from NCDOT\Lidar2007_Clay\LIDAR DEMs.mdb\FocMn3_elevdif';'z:\SAFC_INFO\LIDAR from NCDOT\Lidar2007_Clay\LIDAR DEMs.mdb\FocDif5_FC';'z:\SAFC_INFO\LIDAR from NCDOT\Lidar2007_Clay\LIDAR DEMs.mdb\FocDif3_FC';'z:\SAFC_INFO\LIDAR from NCDOT\Lidar2007_Clay\LIDAR DEMs.mdb\nhdfl_FC_strmslp';'z:\SAFC_INFO\LIDAR from NCDOT\Lidar2007_Clay\LIDAR DEMs.mdb\Catch_NLCD41_2';'z:\SAFC_INFO\LIDAR from NCDOT\Lidar2007_Clay\LIDAR DEMs.mdb\Catch_NLCD42_2';'z:\SAFC_INFO\LIDAR from NCDOT\Lidar2007_Clay\LIDAR DEMs.mdb\Catch_NLCD43_2';'z:\SAFC_INFO\LIDAR from NCDOT\Lidar2007_Clay\LIDAR DEMs.mdb\can_FC_focmn';'z:\SAFC_INFO\LIDAR from NCDOT\Lidar2007_Clay\LIDAR DEMs.mdb\can_FC_focmax'", Sample_26_2_1, Samp_26_2_1, "NEAREST")

Figure C-2:

Mapping.py
# Created on: Sun Apr 13 2008 10:14:07 PM
# (generated by ArcGIS/ModelBuilder)

# Import system modules
import sys, string, os, arcgisscripting

# Create the Geoprocessor object
gp = arcgisscripting.create()

# Check out any necessary licenses
gp.CheckOutExtension("spatial")

# Load required toolboxes...
gp.AddToolbox("C:/Program Files/ArcGIS/ArcToolbox/Toolboxes/Spatial Analyst Tools.tbx")
gp.AddToolbox("C:/Program Files/ArcGIS/ArcToolbox/Toolboxes/Conversion Tools.tbx")

# Set the Geoprocessing environment...
gp.XYResolution = ""
gp.scratchWorkspace = "z:\SAFC_INFO\LIDAR from NCDOT\Lidar2007_Clay\LIDAR DEMs.mdb"
gp.MTolerance = ""
gp.compression = "LZ77"
gp.randomGenerator = "0 ACM599"
gp.outputCoordinateSystem = "PROJCS['NAD_1983_StatePlane_North_Carolina_FIPS_3200_Feet', GEOGCS['GCS_North_American_1983', DATUM['D_North_American_1983', SPHEROID['GRS_1980', 6378137.0, 298.25722101], PRIMEM['Greenwich', 0.0], UNIT['Degree', 0.0174532925199433]], PROJECTION['Lambert_Conformal_Conic'], PARAMETER['False_Easting', 2000000.002616666], PARAMETER['False_Northing', 0.0], PARAMETER['Central_Meridian', -79.0], PARAMETER['Standard_Parallel_1', 34.33333333333334], PARAMETER['Standard_Parallel_2', 36.166666666666666], PARAMETER['Latitude_Of_Origin', 33.75]], UNIT['Foot_US', 0.3048006096012192]]"
gp.rasterStatistics = "STATISTICS 1 1"
gp.outputZFlag = "Same As Input"
gp.qualifiedFieldNames = "true"
gp.tileSize = "128 128"
gp.pyramid = "PYRAMIDS -1 NEAREST"
gp.extent = "526222.46774563 511633.278643592 587602.46774563 547433.278643592"
gp.XYTolerance = ""
gp.cellSize = "20"
gp.outputMFlag = "Same As Input"
gp.geographicTransformations = ""
gp.ZResolution = ""
gp.outputZValue = ""
gp.outputZFlag = "Same As Input"
gp.XYTolerance = ""
gp.extent = "526222.46774563 511633.278643592 587602.46774563 547433.278643592"
gp.XYTolerance = ""
gp.cellSize = "20"
gp.outputMFlag = "Same As Input"
gp.geographicTransformations = ""
gp.ZResolution = ""
gp.outputZValue = ""
gp.outputZFlag = "Same As Input"
gp.ZTolerance = ""
gp.extent = "526222.46774563 511633.278643592 587602.46774563 547433.278643592"
gp.XYTolerance = ""
gp.cellSize = "20"
gp.outputMFlag = "Same As Input"
gp.geographicTransformations = ""
gp.ZResolution = ""
gp.outputZValue = ""
gp.outputZFlag = "Same As Input"
gp.ZTolerance = ""
gp.extent = "526222.46774563 511633.278643592 587602.46774563 547433.278643592"
gp.XYTolerance = ""
gp.cellSize = "20"
gp.outputMFlag = "Same As Input"
gp.geographicTransformations = ""
gp.ZResolution = ""
gp.outputZValue = ""
gp.outputZFlag = "Same As Input"
gp.ZTolerance = ""
gp.extent = "526222.46774563 511633.278643592 587602.46774563 547433.278643592"
gp.XYTolerance = ""
gp.cellSize = "20"
gp.outputMFlag = "Same As Input"
gp.geographicTransformations = ""
gp.ZResolution = ""
gp.outputZValue = ""
gp.outputZFlag = "Same As Input"
gp.ZTolerance = ""
MR3_3ma = "z:\SAFC_INFO\LIDAR from NCDOT\Lidar2007_Clay\LIDAR DEMs.mdb\MR3_3ma"

nhdfl_FC_strmslp__3_ = "nhdfl_FC_strmslp"
FocDif5_FC__7_ = "FocDif5_FC"
Slope2_FC__7_ = "Slope2_FC"
can_FC_focmn__9_ = "can_FC_focmn"

MR3_TerNd4 = "z:\SAFC_INFO\LIDAR from NCDOT\Lidar2007_Clay\LIDAR DEMs.mdb\MR3_TerNd4"
ExMR3__4_ = "z:\SAFC_INFO\LIDAR from NCDOT\Lidar2007_Clay\LIDAR DEMs.mdb\ExMR3__4_"
nhdfl_FC_strmslp__4_ = "nhdfl_FC_strmslp"
FocDif5_FC__8_ = "FocDif5_FC"
Slope2_FC__8_ = "Slope2_FC"
can_FC_focmn__10_ = "can_FC_focmn"

MR3_TerNd5 = "z:\SAFC_INFO\LIDAR from NCDOT\Lidar2007_Clay\LIDAR DEMs.mdb\MR3_TerNd5"
ExMR3__5_ = "z:\SAFC_INFO\LIDAR from NCDOT\Lidar2007_Clay\LIDAR DEMs.mdb\ExMR3__5_"
nhdfl_FC_strmslp__5_ = "nhdfl_FC_strmslp"
FocMn5_elevdif__3_ = "FocMn5_elevdif"
FocMn5_elevdif__4_ = "FocMn5_elevdif"
FocMn5_elevdif__5_ = "FocMn5_elevdif"

MR3_All_2 = "z:\SAFC_INFO\LIDAR from NCDOT\Lidar2007_Clay\LIDAR DEMs.mdb\MR3_All_2"
Tab_MR4All = "z:\SAFC_INFO\LIDAR from NCDOT\Lidar2007_Clay\LIDAR DEMs.mdb\Tab_MR4All"
Sample_26 = "Sample_26"
Sample_26_2_1 = "Sample_26_2_1"
Tab_MR4All2 = "z:\SAFC_INFO\LIDAR from NCDOT\Lidar2007_Clay\LIDAR DEMs.mdb\Tab_MR4All2"

FocMn5_elevdif = "z:\SAFC_INFO\LIDAR from NCDOT\Lidar2007_Clay\LIDAR DEMs.mdb\FocMn5_elevdif"

FocMn5_elevdif__3_ = "FocMn5_elevdif"
can_FC_focmn__10_ = "can_FC_focmn"

FocMn5_elevdif__5_ = "FocMn5_elevdif"

# Process: Single Output Map Algebra...
gp.SingleOutputMapAlgebra_sa("CON(((( FocDif5_FC > 24.5 ) AND (can_FC_focmn < 96.8 ) AND (Slope2_FC < 11.1))),1,0)", MR4_TerNd1, "FocDif5_FC;Slope2_FC;can_FC_focmn")

# Process: Polyline to Raster...
gp.PolylineToRaster_conversion(nhdfl_study, "RESOLUTION", nhdfl_studyRa, "MAXIMUM_LENGTH", "NONE", "20")

# Process: Reclassify...
gp.Reclassify_sa(nhdfl_studyRa, "Value", "0 1", nhdfl_SA_strms, "DATA")

# Process: Extract by Mask...
gp.ExtractByMask_sa(MR4_TerNd1, nhdfl_SA_strms, ExMR4_1)

# Process: Single Output Map Algebra (2)...
gp.SingleOutputMapAlgebra_sa("CON(((( FocDif5_FC > 24.5 ) AND (can_FC_focmn < 96.8 ) AND (Slope2_FC > 11.1) AND (can_FC_focmn < 84.5))),1,0)", MR4_TerNd2, "FocDif5_FC;Slope2_FC;can_FC_focmn")

# Process: Extract by Mask (2)...
gp.ExtractByMask_sa(MR4_TerNd2, nhdfl_SA_strms, ExMR4_2)

# Process: Single Output Map Algebra (3)...
gp.SingleOutputMapAlgebra_sa("CON(((( FocDif5_FC > 24.5 ) AND (can_FC_focmn < 96.8 ) AND (Slope2_FC > 11.1) AND (can_FC_focmn > 84.5))),(1,0)", MR4_TerNd3, "FocDif5_FC;Slope2_FC;can_FC_focmn")

# Process: Extract by Mask (3)...
gp.ExtractByMask_sa(MR4_TerNd3, nhdfl_SA_strms, ExMR4_3)

# Process: Single Output Map Algebra (4)...
gp.SingleOutputMapAlgebra_sa("CON(((((( FocDif5_FC > 24.5 ) AND (can_FC_focmn < 96.8 ) AND (Slope2_FC > 11.1) AND (can_FC_focmn > 84.5)) AND (Slope2_FC > 16.3) AND (can_FC_focmn < 95) AND (FocMn5_elevdif > 2.4))))),(1,0)", MR4_TerNd4, "FocDif5_FC;Slope2_FC;can_FC_focmn;FocMn5_elevdif;FocDif3_FC")

# Process: Extract by Mask (4)...
gp.ExtractByMask_sa(MR4_TerNd4, nhdfl_SA_strms, ExMR4_4)

# Process: Single Output Map Algebra (5)...
gp.SingleOutputMapAlgebra_sa("CON((((((( FocDif5_FC > 24.5 ) AND (can_FC_focmn < 96.8 ) AND (Slope2_FC > 11.1) AND (can_FC_focmn > 84.5)) AND (Slope2_FC > 16.3) AND (can_FC_focmn < 95) AND (FocDif5_FC > 21.5 AND FocDif3_FC < 25.5) AND (FocMn5_elevdif > 2.4))))),(1,0)", MR4_TerNd5, "FocDif5_FC;Slope2_FC;can_FC_focmn;FocMn5_elevdif;FocDif3_FC")
gp.ExtractByMask_sa(MR4_TerNd5, nhdfl_SA_strms, ExMR4_5)

gp.SingleOutputMapAlgebra_sa("ExMR4_1 + ExMR4_2 + ExMR4_3 + ExMR4_4 + ExMR4_5", MR4_All, 
"'z:\SAFC_INFO\LIDAR from NCDOT\Lidar2007_Clay\LIDAR DEMs.mdb\ExMR4_1';
'z:\SAFC_INFO\LIDAR from NCDOT\Lidar2007_Clay\LIDAR DEMs.mdb\ExMR4_2';
'z:\SAFC_INFO\LIDAR from NCDOT\Lidar2007_Clay\LIDAR DEMs.mdb\ExMR4_3';
'z:\SAFC_INFO\LIDAR from NCDOT\Lidar2007_Clay\LIDAR DEMs.mdb\ExMR4_4';
'z:\SAFC_INFO\LIDAR from NCDOT\Lidar2007_Clay\LIDAR DEMs.mdb\ExMR4_5'")

gp.SingleOutputMapAlgebra_sa("CON(((nhdfl_FC_strmslp > 0.15761) AND (can_FC_focmn < 90.6667))", MR3_TerNd1, 
"can_FC_focmn;nhdfl_FC_strmslp")
APPENDIX D: SELECTED R CODE FOR CART MODEL

setwd('z:/MP/MP_Tables/')
csv='z:/MP/MP_Tables/Samp26_21908_soilstoo.csv'
data26 = read.csv(csv)
summary(data26)  #summary stats for variables in columns
dim(data26)

##how to subset rows from http://www.ats.ucla.edu/stat/r/modules/subsetting.htm
##ex. write.50 <- subset(hsb2.small, write > 50)
##this section is good for getting my zeros and ones subset
zeros <- subset(data26, block==0)
summary(zeros)
dim(zeros)
ones <- subset(data26, block==1)
dim(ones)

##from Ben, help on the sampling of 0s issue
irows=sample(1:nrow(zeros),100,replace=False) #samples the row index from my subset of 0s
zeros.irows=data26[irows,]  #extract data from all columns using that index (fetch data)

##from Dean, original code to bind data into new dataset (after subsetting and sampling from 0s) with which i proceed to CART
new.data26 <- rbind(ones, zeros.irows)
dim(new.data26)
summary(new.data26)

##CART START
library(tree)  #NEED TO LOAD FIRST!!!
help(package=tree)
new.data26tree <- tree(as.factor(new.data26$block) ~ slp + mn5ed + mn3ed + focdif5 + focdif3 + strslp + ccmn + ccmax, data=new.data26, method="class")
plot(new.data26tree)
text(new.data26tree, cex=0.8)
summary(new.data26tree)  #returns the model call, lists variables actually used, reports residual deviance(regression trees) and overall misclass rate
print(new.data26tree)  #returns tree details

#TREE PRUNING
help(prune.tree)
new.data26tree.prune <- prune.misclass(new.data26tree)  # prune tree using misclassifications
print(new.data26tree.prune)  # prints pruned tree to screen
plot(new.data26tree.prune)  # plots the misclassification rate as a function of tree size

#CROSS VALIDATION - procedure randomly partitions data into 10 subsets, then estimates a tree using each 90% partition
#and predicts group membership for the 10% of the data withheld from the partition. repeats until all samples independently classified,
#must specify the kind of function to prune tree so use method="misclass"
new.data26tree.cv<- cv.tree(new.data26tree, FUN=prune.tree, method="misclass", K=20)
par(mfrow=c(2,1))
plot(new.data26tree.prune)
plot(new.data26tree.cv)

#CLASSIFICATION SUCCESS AND CONFUSION MATRIX
new.data26.pred <- predict(tree(new.data26tree, type="class")
#suppose to give table with rows are predicted group membership and
#columns are actual values
table(new.data26.pred, new.data26$block)