The Importance of Scale: Assessing and Predicting Brook Trout Status in its Southern Native Range

Mark Hudy[1]
U.S. Forest Service, Fish and Aquatic Ecology Unit, James Madison University, Harrisonburg, Virginia 22807

Keith Nislow
U.S. Forest Service, Northern Research Station, University of Massachusetts, Amherst, Massachusetts, 1003

Eric P. Smith
Virginia Tech, Department of Statistics, Blacksburg, Virginia 24061

Arthur R. Cooper and Dana M. Infante
Michigan State University, Department of Fisheries and Wildlife, Lansing Michigan, 48824

[1] Current address: *U.S. Geological Survey, 12201 Sunrise Valley Drive, Reston, Virginia 20192*

Abstract. -- Occupancy models are of increasing interest to managers and natural resource decision makers. Assessment of status and trends, as well as the specific drivers influencing occupancy, both may change as a function of scale, and analyses conducted at multiple scales can help identify important mechanisms leading to changes in distributions. We analyzed extensive fine-scale occupancy data across the southern historic range of the brook trout, *Salvelinus fontinalis* to determine which landscape metrics and thresholds were useful in predicting brook trout presence across three relevant spatial scales and how brook trout occupancy varied by scale. Percentage occupancy declined markedly with increased spatial resolution, as 52% of watersheds (HUC10) but only 32% of subwatersheds (HUC12) and 14% of catchments (HUC14) were occupied. Across all three scales, habitats which were exclusively occupied by native brook trout (without non-native trout) were rare (<10%). CART models using GIS-derived landscape predictor variables were developed for three classification cases: Case 1:(brook trout; no brook trout), Case 2 (brook trout; non-native trout only; no trout), and Case 3 (brook trout only; brook and non-native trout; non-native trout only and no trout). Model results were sensitive to both scale and the number of classification categories with respect to classification accuracy, variable selection and variable threshold values. Classification accuracy tended to be lowest at the finest (catchment) scale potentially reflecting stochastic population processes and barriers to movement. Classification rates for the overall models were: Case 1: Watershed (80.19%); Subwatershed (85.06%); Catchment (71.13%); Case 2: Watershed (69.31%); Subwatershed (68.72%); Catchment (57.38%); Case 3: Watershed (58.91%); Subwatershed (59.83%); Catchment (47.59%). Our multiscale approach revealed soil permeability (positive) and atmospheric pollution (negative) to be important predictors. The predicted occupancy and observed status of brook trout appear to be influenced by the scale the data are collected and reported.

Introduction

Maintaining sustainable and resilient wild trout populations depends on the availability of a distinct set of habitat conditions that are threatened by a wide range of anthropogenic factors operating at multiple spatial and temporal scales. These concerns, along with apparent declines in distribution and abundance in some systems, have led to calls for systematic analysis of status and trends, particularly at large landscape scales (Rieman et al. 1997; Thurow et al. 1997; Hudy et al. 2008). In addition, given likely changes in habitat suitability in the near- and longer-term, managers need tools, which will help them identify, conserve, and restore priority populations and habitats. Occupancy models, which identify habitat and landscape variables that are associated with occurrence, can be useful for these purposes, and are increasingly used in the context of wild trout conservation (Wagner et al. 2013).

Brook trout, *Salvelinus fontinalis*, provide an excellent example of these issues. Native populations, particularly in the southern portion of their range, are under threat from landscape change (deforestation and development), non-native species (particularly naturalized rainbow trout and brown trout), atmospheric pollution, and most recently, regional climate warming. These concerns have led to the creation of the Eastern Brook Trout Joint Venture (EBTJV) (EBTJV 2006) a multiagency effort to document status and trends over the entire historic range. Previous work accomplished under the EBTJV produced an initial status assessment (Hudy et al. 2008). This assessment was conducted at relatively coarse resolution (subwatershed (HUC12)) and confirmed that populations appeared to be at risk and declining, and that an initial predictive model identified forest cover as an important determinant of brook trout occurrence. However, the determination of status and the identification of critical habitat factors are highly dependent on the spatial scale of analysis. For brook trout, these issues are particularly relevant. Many populations exist in small isolated, headwater patches and brook trout presence in some subwatersheds may only be represented by one or a few of these small populations, which are potentially highly vulnerable to extirpation. Similarly, at these fine spatial scales, the factors that best account for brook trout presence may change, with consequent implications for management actions. To address these issues, in this study we used fine-scale (catchment HUC14) occupancy data, along with an expanded list of GIS-based landscape variables, to assess status of both native and non-native trout and the factors contributing to presence or absence of these species over the southern historic range of the brook trout. Our goal was to refine our assessment of status, and to increase our ability to identify priority habitats and effective management strategies.

Methods

Study area, assessment scale and classifications of reproducing trout

Through over 100 years of sampling efforts, an extensive and fine scale database of cold-water habitats supporting naturally reproducing trout has been developed within the southern historic range of the brook trout (Pennsylvania, New Jersey, Maryland, West Virginia, Virginia, North Carolina, Tennessee, Georgia, South Carolina)(Hudy et al. 2008). These states are unique within the EBTJV in that they have a census of the downstream extant of reproducing trout. Detection probabilities are unknown for the majority of the datasets but ranged from 89 to 99 % in Pennsylvania streams (Wagner et al. 2013).

We created a 50-km buffer zone around a 1969 map of the species' native distribution in the study area (developed from fish collections and personal communications with fisheries experts; MacCrimmon and Campbell 1969; Hudy et al. 2008) and classified all 808 watersheds (HUC10); 3,804 subwatersheds (HUC12) and 132,321 catchments (HUC14) that were situated wholly or partially within the region and buffer zone according to occupancy of reproducing populations of native brook trout, and non-native rainbow trout *Oncorhynchus mykiss* and brown trout *Salmo trutta* . A total of 107 sub-basins (HUC 8) were within the study area but we did not include these in this analysis because of the small sample sizes in some of the wild trout classes we were trying to predict.

There were 8 possible classifications (Table 1). We grouped these classifications and developed models to predict wild trout status for 3 different cases of interest to biologists participating in the EBTJV. Case 1 models predicted two categories either brook trout (defined as watershed units classified as 1, 2, 3 or 4) or no brook trout (defined as watershed units classified as 5,6,7,or 8). Case 2 models predicted three categories; brook trout (defined as watershed units classified as 1, 2, 3 or 4), non-native trout (defined as watershed units classified as 5, 6, or 7) and no trout (defined as watershed units classified as 8). Case 3 models predicted 4 categories; allopatric brook trout (defined as watershed units classified as 1), sympatric brook trout (defined as watershed units classified as 2, 3 or 4), non-native trout (defined as watershed units classified as 5, 6, or 7) and no trout (defined as watershed units classified as 8). Models were developed for all three cases for all three-watershed scales for a total of 9 models.

Candidate metrics, metric screening, metric calculations

We calculated and evaluated 85 candidate landscape metrics (Thieling 2006; Esselman 2011) and these metrics were used instead of site-specific variables (Moyle and Randall 1998) (Table 2). Each metric was summarized by the relevant watershed (HUC10); subwatershed (HUC12) or 4 catchment (HUC14) categories (LC= local catchment; NC= network catchment (sum of all upstream catchments); LB = local catchment buffer (sum of 100 m buffer around NHD+ stream layer in the local catchment) and NB= network buffer (sum of 100 m buffer of all upstream NHD+ streams). Landscape metrics can provide an indicator of watershed health when many anthropogenic factors potentially contribute to a problem, and such metrics can assist in identification of key limiting factors (Marschall 1996; McCormick et al. 2001).

Model development

The data were analyzed using CART version 7.0 (http://www.salford-systems.com., Brieman et al., 1984) using classification trees. Classification trees are popular as they are relative easy to understand and are nonparametric methods. A classification tree is a series of binary splits that results in nodes and branches that produces a prediction model for a class variable. At each node a split of the data is made based on a threshold value for one of the independent variables. Each series of nodes leads to a terminal node, provides a classification rule and is associated with independent variables and associated thresholds. The decision as to what variable is used and the threshold is based on a measure of change in a classification metric. For this analysis, the Gini measure of "purity" of the classification was used. A tree grows until no improvements may be made and often results in overfitting of a model. Part of the goal is to produce a simple tree. To do this we used ten-fold crossvalidation and used a one-standard error rule. With crossvalidation, the data are split ten times into a group consisting of 90% of the data and 10% of the data. The model is fit using the 90% group and evaluated using the 10% group. The process is repeated 10 times using a different subset each time. A "cost" or classification function is used to measure the overall quality of the tree, and crossvalidation results in an estimate of the variance of the measure. A simple tree is one that has a cost that is similar to the best tree (i.e. within one standard error of the best tree) but has fewer nodes. Following selection of the tree variable importance measures are calculated along with prediction rates. Variable importance measures how much of the reduction in cost is associated with a particular variable. CART allows use of all variables (including those never used in the tree) to be included or the use of variables associated with splitting of the tree. Variable importance measures are scaled so the greatest value is 100. We used just the variables associated with splits in the calculations. Prediction rates were also calculated on test data to provide a conservative estimate of prediction success. The catchment data have a large number of observations and this lead to difficulties with the memory for the CART analysis. To simplify the analysis and to use all the observations, we first used the R package rpart to identify important variables and used a subset of the variables in the analysis.

Results and Discussion

Using extensive, fine scale, catchment occupancy data we found that brook trout currently occupy 51.61% of the watersheds; 31.91% of the subwatersheds and 13.97% of the catchments in the southeastern United States within their historic native range (Table 3). In the original EBTJV assessment the analysis was limited to subwatersheds because the lack of extensive fine scale data in the northeastern United States (Hudy et al. 2008). These patterns of occupancy have important implications for status and vulnerability assessment. Although there is some uncertainty with respect to historic occupancy rates at the finest (catchment) scale, our observation that < 14% of catchments are currently occupied suggest a high level of vulnerability to current and future threats, with brook trout restricted to one or two small populations in isolated headwater habitats. While current and historic data collected from EBTJV members clearly show that unoccupied watersheds and subwatersheds historically had brook trout and now have been extirpated (Hudy et al. 2008) our results suggest that these analyses may underestimate risk. Re-evaluating subwatersheds at the catchment level has improved identification of these at risk subwatersheds and helped set more informed priorities for restoration and conservation efforts. For example approximately 19 % of the catchments in North Carolina lost brook trout since the 2005 assessment (Hudy et al. 2008). In other cases, previously unoccupied catchments were recolonized or restored (less than 2%). Based on this new analysis agencies managing the EBTJV and the Chesapeake Bay Executive Order have adopted catchment level metrics for monitoring status and trends.

The majority of brook trout populations throughout the southern range co-occur with non-native trout (rainbow and/or brown trout (watersheds: 7.55% allopatric; 44.06% sympatric; subwatersheds: 9.75% allopatric; 22.75% sympatric; catchments: 7.81% allopatric; 6.17% sympatric). While the establishment of non-native trout in brook trout watersheds has been previously addressed in a number of studies (Kelly et al. 1980; Larson and Moore 1985; Wagner et al. 2013), ours is the first to broadly document these patterns of co-occurrence across multiple scales. The data suggest that even catchments that currently contain only native brook trout are vulnerable to invasion from non-native trout established within their subwatersheds. These invaded populations may be at increased risk, as some studies have found negative impacts of non-native trout on wild brook trout (Kelly et al. 1980; Wagner et al. 2013) and non-native trout may also be better able to persist in anthropogenically altered habitats (Wagner et al. 2013). Monitoring efforts need to focus on rates of non-native establishment at the catchment scale and the fate and status of brook trout populations currently in sympatry, particularly as compared to the few remaining allopatric populations. If negative impacts are detected, this would place particular emphasis on keeping those few subwatersheds that currently contain only brook trout free of non-native species.

Using CART models to predict occupancy yielded relatively high classification accuracy in most cases. Model results were sensitive to both scale and the number of classification categories with respect to classification accuracy, variable selection and variable threshold values (Table 3). Overall prediction rates were: Case 1 (watershed 80.19%; subwatershed 85.06%; catchment 71.13%); Case 2 were (watershed 69.31%; subwatershed 68.72%; catchment 57.38%) and Case 3 (watershed 58.91%; subwatershed 59.83%; catchment 47.59%)(Table 3). For all three cases, classification accuracy was lowest at the catchment scale, potentially reflecting stochastic population processes and barriers to movement, which would complicate predication based on landscape variables. Our best predictive models were for Case 1 at the subwatershed scale. A total of 39 of the original 85 metrics were used at least once in the 9 different models (Table 4). The metric of soil permeability was the most important variable for watershed and subwatershed models but was not important at the catchment scale (Table 4). Many of the important prediction metrics are fixed (i.e. elevation) but the majority can be influenced both short term and long term by management actions (i.e. soil permeability; canopy cover; land use; road density, $S0_4$ and $N0_3$ deposition). These models are therefore useful in targeting actions (increasing canopy cover, decommissioning roads, liming watersheds) that would increase brook trout population resilience.

Most of the data used here were provided by state and federal agencies and had not been published or peer reviewed. Despite the criteria developed for status classification, there remains some element of subjectivity in the assessment of status and occupancy. It was impossible to generate a comprehensive review without such data (Reiman et al. 1997). We attempted to limit errors, reduce subjectivity, and provide consistency in data by using consistency rules and data standards (quality and age); developing broad classification categories; and employing standard, validated procedures in consulting experts (Hudy et al. 2008). Moving forward, continued improvement and standardization of decision rules and data collection protocols, along with increasing understanding of the factors limiting brook trout population and incorporation of these factors into new models, will further increase our ability to assess status, predict trends, and target management actions.

Acknowledgements:

We thank the members of the EBTJV and the National Fish Habit Board Science and Data committee and Inland Assessment Team for providing their latest data.

References

Breiman, L., J. Friedman, R. Olshen, and C. Stone. Classification and Regression Trees. Pacific Grove:Wadsworth, 1984. http://www.salford-systems.com.

EBTJV (Eastern Brook Trout Joint Venture). 2006. About the Eastern Brook Trout Venture. Available: www. easternbrooktrout.org. (December 2006).

Esselman,P.T., D.M. Infante, L. Wang, D.Wu, A.R. Cooper and W.W. Taylor. 2011. An Index of Cumulative Disturbance to River Fish Habitats of the Conterminus United States from Landscape Anthropogenic Activities. Ecological Restoration 29:1-2

Hudy, M., T. M. Thieling, N. Gillespie, and E. P. Smith. 2008. Distribution, status, and perturbations to brook trout within the eastern United States. North American Journal of Fisheries Management 28:1069 -1085.

Kelly, G. A., J. S. Griffith, and R. D. Jones. 1980. Changes in distribution of trout in Great Smoky Mountains National Park, 1900–1970. U.S. Fish and Wildlife Service Technical Paper 102.

Larson, G. L., and S. E. Moore. 1985. Encroachment of exotic rainbow trout into stream populations of native brook trout in the southern Appalachian Mountains. Transactions of the American Fisheries Society 114:195–203.

MacCrimmon, H. R., and J. S. Campbell. 1969. World distribution of brook trout, Salvelinus fontinalis. Journal of the Fisheries Research Board of Canada 26:1699– 1725.

Marschall, E. A., and L. B. Crowder. 1996. Assessing population responses to multiple anthropogenic effects: a case study with brook trout. Ecological Applications 6:152–167.

McCormick, F. H., R. M. Hughes, P. R. Kaufmann, D. V. Peck, J. L. Stoddard, and A. T. Herlihy. 2001. Development of an index of biotic integrity for the mid-Atlantic Highlands region. Transactions of the American Fisheries Society 130:857–877.

Moyle, P. B., and P. J. Randall. 1998. Evaluating the biotic integrity of watersheds in the Sierra Nevada, California. Conservation Biology 12:1318–1326.

Reiman, B. E., D. C. Lee, and R. F. Thurow. 1997. Distribution, status, and likely future trends of bull trout within the Columbia River and Klamath River basins, North American Journal of Fisheries Management 17:1111–1125.

Thieling, T. M. 2006. Assessment and predictive model for brook trout (Salvelinus fontinalis) population status in the eastern United States. Master's thesis. James Madison University, Harrisonburg, Virginia.

Thurow, R. F., D. C. Lee, and B. E. Reiman. 1997. Distribution and status of seven native salmonids in the interior Columbia River basin and portions of the Klamath River and Great basins. North American Journal of Fisheries Management 17:1094–1110.

Wagner, T., J.T. Deweber, J. Detar and J.A. Sweka. 2013. Landscape-Scale Evaluation of Asymmetric Interactions between Brown Trout and Brook Trout Using Two-Species Occupancy Models. Transactions of the American Fisheries Society 142:353-361.

Table 1. Eight wild trout occupancy categories used to classify 808 watersheds; 3,804 subwatersheds and 132,321 catchments. 1 = present; 0 = absent. Data collected from EBTJV members (Hudy et al. 2008).

Occupancy classifications	**Brook trout**	**Rainbow Trout**	**Brown Trout**
1.	1	0	0
2.	1	1	0
3.	1	0	1
4.	1	1	1
5.	0	0	1
6.	0	1	0
7.	0	1	1
8.	0	0	0

Table 2. Candidate metrics summarized by type, attribute base name, description, units, source, dataset, scale/resolution, date and website. All metrics used for developing predicative regression tree models. Metrics summarized by relevant scale of analysis (watershed, subwatershed, catchment).

Table 3. Occupancy and classification rates (crossvalidation) for watersheds (HUC10); subwatersheds (HUC12) and catchments (HUC14) for three cases of trout classification. Case 1: brook trout, no brook trout; Case 2: brook trout, non-native trout, no trout; Case 3: allopatric brook trout, sympatric brook trout, non-native trout, no trout).

Table 4. Variable importance for predicting brook trout (Case 1: brook trout, no brook trout; Case 2: brook trout, non-native trout, no trout; Case 3: allopatric brook trout, sympatric brook trout, non-native trout, no trout) for three scales (watershed (HUC 10); subwatershed (HUC 12); catchment (HUC 14)) using CART regression tree models. Each metric summarized by the relevant scale of analysis (watershed; subwatershed; catchment). Catchment data summarized by 4 methods for each metric (LC = metric summarized for entire area of local catchment; LB = metric summarized by 100 m buffer of the NHD+ stream layer within the local catchment; NC = metric summarized by the area of the stream network (local catchment plus all upstream catchments).

Attribute Base Name	Description	Units	Scale/Resolution	Currentness
Surficial Lithology, USGS[1]				
LITHO_1	Carbonate Residual Material	proportion	1 km	2004
LITHO_3	Non-Carbonate Residual Material	proportion	1 km	2004
LITHO_5	Silicic Residual Material	proportion	1 km	2004
LITHO_8	Colluvial Sediment	proportion	1 km	2004
LITHO_9	Glacial Till, Clayey	proportion	1 km	2004
LITHO_10	Glacial Till, Loamy	proportion	1 km	2004
LITHO_11	Glacial Till, Coarse-Textured	proportion	1 km	2004
LITHO_12	Glacial Outwash and Glacial Lake Sediment, Coarse-Textured	proportion	1 km	2004
LITHO_13	Glacial Lake Sediment, Fine-Textured	proportion	1 km	2004
LITHO_14	Glacial Outwash and Glacial Lake Sediment, Coarse-Textured	proportion	1 km	2004
LITHO_15	Hydric, Peat and Muck	proportion	1 km	2004
LITHO_16	Eolian Sediment, Coarse-Textured (Sand Dunes)	proportion	1 km	2004
LITHO_17	Eolian Sediment, Fine-Textured (Glacial Loess)	proportion	1 km	2004
LITHO_19	Alluvium and Fine-Textured Coastal Zone Sediment	proportion	1 km	2004
LITHO_20	Coastal Zone Sediment, Coarse-Textured	proportion	1 km	2004
LITHO_22	Water	proportion	1 km	2004
2001 National Land Cover Database, MRLC[2]				
NLCD2001_11	Open Water	proportion	30 m	2001
NLCD2001_21	Developed, Open Space	proportion	30 m	2001
NLCD2001_22	Developed, Low Intensity	proportion	30 m	2001
NLCD2001_23	Developed, Medium Intensity	proportion	30 m	2001
NLCD2001_24	Developed, High Intensity	proportion	30 m	2001
NLCD2001_31	Barren Land (Rock/Sand/Clay)	proportion	30 m	2001
NLCD2001_41	Deciduous Forest	proportion	30 m	2001
NLCD2001_42	Evergreen Forest	proportion	30 m	2001
NLCD2001_43	Mixed Forest	proportion	30 m	2001
NLCD2001_52	Shrub/Scrub	proportion	30 m	2001
NLCD2001_71	Grassland/Herbaceous	proportion	30 m	2001
NLCD2001_81	Pasture/Hay	proportion	30 m	2001
NLCD2001_82	Cultivated Crops	proportion	30 m	2001
NLCD2001_90	Woody Wetlands	proportion	30 m	2001
NLCD2001_95	Emergent Herbaceous Wetlands	proportion	30 m	2001
2006 National Land Cover Database, MRLC[2]				
NLCD2006_11	Open Water	proportion	30 m	2006
NLCD2006_21	Developed, Open Space	proportion	30 m	2006
NLCD2006_22	Developed, Low Intensity	proportion	30 m	2006
NLCD2006_23	Developed, Medium Intensity	proportion	30 m	2006
NLCD2006_24	Developed, High Intensity	proportion	30 m	2006
NLCD2006_31	Barren Land (Rock/Sand/Clay)	proportion	30 m	2006
NLCD2006_41	Deciduous Forest	proportion	30 m	2006
NLCD2006_42	Evergreen Forest	proportion	30 m	2006
NLCD2006_43	Mixed Forest	proportion	30 m	2006
NLCD2006_52	Shrub/Scrub	proportion	30 m	2006
NLCD2006_71	Grassland/Herbaceous	proportion	30 m	2006
NLCD2006_81	Pasture/Hay	proportion	30 m	2006
NLCD2006_82	Cultivated Crops	proportion	30 m	2006
NLCD2006_90	Woody Wetlands	proportion	30 m	2006
NLCD2006_95	Emergent Herbaceous Wetlands	proportion	30 m	2006
Base-flow index grid for the conterminous United States, USGS[3]				
BFI	Baseflow Index (% groundwater contribution to baseflow)	percentage	1 km	2003
Atmospheric Deposition Isopleth Maps, National Atmospheric Deposition Program[4]				
DEPNO3_MEAN	Annual NO3 wet deposition averaged for 1994 – 2010	kilograms/hectare	2.338 km	1994-2010
DEPNO3_MAX	Maximum NO3 wet deposition value taken from 1994 – 2010 a	kilograms/hectare	2.338 km	1994-2010
DEPNO3_MAXALL	Maximum NO3 wet deposition value taken across entire time span (1994 – 2010) b	kilograms/hectare	2.338 km	1994-2010
DEPSO4_MEAN	Annual SO4 wet deposition averaged for 1994 – 2010	kilograms/hectare	2.338 km	1994-2010
DEPSO4_MAX	Maximum SO4 wet deposition value taken from 1994 – 2010 a	kilograms/hectare	2.338 km	1994-2010
DEPSO4_MAXALL	Maximum SO4 wet deposition value taken across entire time span (1994 – 2010) b	kilograms/hectare	2.338 km	1994-2010
National Elevation Dataset (acquired from NHDPlus v.1), USGS[5]				
ELEV_MAX	Maximum catchment elevation	centimeters	30 m	2005
ELEV_MEAN	Average catchment elevation	centimeters	30 m	2005
ELEV_MIN	Minimum catchment elevation	centimeters	30 m	2005
2010 U.S. Census, US Census Bureau[6]				
POPDENS	Population density	#/km^2	1:100,000 c	2010
HUDENS	Housing unit density	#/km^2	1:100,000 c	2010
2001 National Land Cover Database, MRLC[2]				
NLCD2001_FC	Tree canopy coverage	percentage	30 m	2001
NLCD2001_IS	Impervious surface coverage	percentage	30 m	2001
2006 National Land Cover Database, MRLC[2]				
NLCD2006_IS	Impervious surface coverage	percentage	30 m	2006
STATSGO, USGS[7]				
SOILPERM	Average soil permeability rates	inches/hour * 100	1:250,000	1995
U.S. Climate Normals (1981-2010), PRISM Climate Group[8]				
PPT2010	Average annual precipitation (1981-2010)	millimeters * 100	800 m	1981-2010
TMAX2010	Average maximum annual air temperature (1981-2010)	degrees C *100	800 m	1981-2010
TMEAN2010	Average annual air temperature (1981-2010)	degrees C *100	800 m	1981-2010
TMIN2010	Average minimum annual air temperature (1981-2010)	degrees C *100	800 m	1981-2010
N/A	Zone area		N/A	
AREASQKM	Zone area	km^2	N/A	N/A
2012 National Anthropogenic Barrier Dataset, USGS				
Dam_Dens	Dam density	#/km of stream network	N/A	2012
Dam_Count	2009 National Inventory of Dams count	count	N/A	2009
Active Mines and Mineral Processing Plants, USGS[9]				
Mine_Dens	Mine/mineral processing plant density	#/km of stream network	N/A	2003
CERC_Dens	EPA Superfund National Priority (CERC) site density	#/km of stream network	N/A	2007
PCS_Dens	EPA National Pollutant Discharge Elimination System site density	#/km of stream network	N/A	2007
TRI_Dens	EPA Toxics Release Inventory site density	#/km of stream network	N/A	2007
Roadcr_Dens	Road/stream network crossing density	#/km of stream network	1:100,000	2011
Road_Dens	Road length density	km of road/km of stream	1:100,000	2011
Active Mines and Mineral Processing Plants, USGS[9]				
Mine_Count	Mine/mineral processing plant count	count	N/A	2003
Geodata Shapefile, US EPA[10]				
CERC_Count	EPA Superfund National Priority (CERC) site count	count	N/A	2007
PCS_Count	EPA National Pollutant Discharge Elimination System site density	count	N/A	2007
TRI_Count	EPA Toxics Release Inventory site count	count	N/A	2007
TIGER 2011 Roads/NHDPlus v.1, US Census Bureau[11]				
Roadcr_Count	Road/stream network crossing count	count	1:100,000	2011
Roadcr_Dens	Road/stream network crossing density	#/km of stream network	1:100,000	2011
Road_Dens	Road length density	km of road/km of stream	1:100,000	2011
Roadlen_km	Total road length	km	1:100,000	2011
NHDPlus v.1, US EPA/USGS[5]				
Netlen_km	Total NHDPlus v.1 hydrography network length d	km	1:100,000	2005
Strlen_km	Total NHDPlus v.1 stream network length e	km	1:100,000	2005

1. http://rmgsc.cr.usgs.gov/ecosystems/usa.shtml
2. http://www.mrlc.gov/
3. http://water.usgs.gov/GIS/metadata/usgswrd/XML/bfi48grd.xml
4. http://nadp.sws.uiuc.edu/
5. http://www.horizon-systems.com/nhdplus/
6. http://www.census.gov/geo/www/tiger/tgrshp2010/tgrshp2010.html
7. http://water.usgs.gov/GIS/metadata/usgswrd/XML/ussoils.xml
8. http://www.prism.oregonstate.edu/index.phtml
9. http://mrdata.usgs.gov/mineral-resources/active-mines.html
10. http://www.epa.gov/geospatial/
11. http://www.census.gov/geo/www/tiger/tgrshp2011/tgrshp2011.html
12. http://geo.usace.army.mil/pgis/f?p=397:1:0

a = averaged grid (maximum value within spatial zone)
b = on a grid cell basis (average maximum within spatial zone using highest value for each grid cell across all years)
c = base unit is census block group
d = includes artificial paths through waterbodies
e = includes only reaches identified as "streams/rivers"
for the national NFHAP assessment

Watershed unit	HUC10	HUC12		Catchment	
	Correct Classification Rate (crossvalidation)		Correct Classification Rate (crossvalidation)		Correct Classification Rate (crossvalidation)
CASE 1 HUC10 (n) (% of total)	Case1 HUC10	CASE 1 HUC12 (n) (% of total)	Case1 HUC12	CASE 1 Catchment (n) (% of total)	Case1 Catchment
brook trout (417) (51.61%)	83.45 (75.06)	brook trout (1,214) (31.91%)	87.81 (80.56)	brook trout (18,491)(13.97%)	80.16 (73.26)
no brook trout (391) (48.39%)	76.73 (72.63)	no brook trout (2,590) (68.09%)	83.78 (79.77)	no brook trout (113,830)(86.03%)	69.97 (69.15)
Overall correct rate (808)	80.19 (73.88)	Overall correct rate (3,804)	85.06 (80.02)	Overall correct rate (132,321)	71.13 (69.72)
Case 2 HUC10 (n) (% of total)	Case2 HUC10	Case 2 HUC12 (n) (% of total)	Case2 HUC12	Case 2 Catchment (n) (% of total)	Case2 Catchment
brook trout (417) (51.61%)	70.02 (63.55)	brook trout (1,214) (31.91%)	71 (68.20)	brook trout (18,491) (13.97%)	60.88 (56.34)
non-native trout (74) (9.16%)	79.73 (60.81)	exotic trout (357) (9.38%)	79.55 (61.06)	exotic trout (9,917)(7.49%)	73.1 (61.55)
no trout (317) (39.23%)	65.93 (60.57)	no trout (2,233) (58.71%)	65.74 (62.70)	no trout (103,913)(78.53%)	55.26 (53.27)
Overall correct rate (808)	69.31 (62.12)	Overall correct rate (3,804)	68.72 (64.30)	Overall correct rate (132,321)	57.38 (54.31)
Case 3 HUC10 (n) (% of total)	Case3 HUC 10	Case 3 HUC12 (n) (% of total)	Case3 HUC12	Case 3 Catchment (n) (% of total)	Case3 Catchment
Allopatric brook trout (61) (7.55%)	72.13 (42.62)	Allopatric brook trout (371) (9.75%)	59.03 (54.45)	Allopatric brook trout (10,329)(7.81%)	47.89 (41.34)
Sympatric brook trout (356) (44.06%)	46.91 (46.63)	Sympatric brook trout (843) (22.16%)	62.16 (51.72)	Sympatric brook trout (8,162)(6.17%)	65.9 (53.30)
non-native trout (74) (9.16%)	72.97 (52.70)	Exotic trout only (357) (9.38%)	71.99 (61.62)	Exotic trout only (9,917)(7.49%)	67.87 (56.70)
no trout (317) (39.23%)	66.56 (63.72)	no trout (2,233) (58.70%)	57.14 (56.29)	no trout (103,913)(78.53%)	44.20 (44.12)
Overall correct rate (808)	58.91 (46.41)	Overall correct rate (3,804)	59.83 (55.59)	Overall correct rate (132,321)	47.59 (45.43)

Watershed Unit	HUC10 Case1	HUC12 Case1	Catchment Case1	HUC10 Case2	HUC12 Case2	Catchment** Case2	HUC10 Case3	HUC12 Case3	Catchment Case3
Metric Variable									
SOILPERM	100	100		100	100		100	100	
NO3 MAX	21.7	47.9	NB=100.00	66.38	11.11		57.84	7.48	
ELEV_Mean	44.89			43.3	86.8	NB=68.57; LB=13,52	22.43	28.18	NB=56.64; LB=56,48
ELEV_MAX		37.88	LB=24,50	53.1	54.65	LB=14,82	38.42	22.93	
NO3_MAXALL			NB=36.42		16.54	LC=100,00; LB=86,38	56.28	14.58	LB=100,00; NC=73,19
DEPOSITION				85.91	10.17	LB=14,46	74.8		NB=69.76
NCLD2001_FC		37.43		37.66	31.47		34.29	32.36	
TMEAN2010					20.69	NC=59.68		95.81	NC=59.48
NLCD2001_71		4.49			22.57		31.81	26.07	
BFI		12.92	LB=84,67; NC=53,90; LC=48.66		46.89			83.11	NC=82.79
TMIN2010		11,35	NC=76.86		74.07	LB=96,67; NC=76,59		18.87	LB=99,71;NC=41.14
ELEV_MIN			NC=63.83		9.4	LB=97,48		7.57	LB=28,08
TMAX2010		8.03	NB=78.29					13.7	NC=30.55
SO4_MAX		3.87	NB= 89,52		43.29			36.67	
PPT2010				33.23	6.43			10.65	
POPDENS		9.28			12.74				
NLCD2001_52		4.62							NB=15.48
ROADCR_DENS	14.81							14.34	
NLCD2001_43				45.24			66.07		
POPDENS		9.28			12.74				
NLCD2001_52		4.62							NB=15.48
TOTALFORESTI		3.99						13.34	
LITHO_8		2.46						12.54	
S04_MEAN			LC=90,29			LB=87,37			
S04_MAXALL			NB=72.51			LB=54,21			
LITHO_11						NB=39.01			NB=33.72
NLCD2006_71	47.01								
LITHO_3									
ROADLEN_KM		1.88							
N03_MEAN			NB=50.59						
LITHO_19					15.27				
NETLEN_KM							52.55		
NLCD2001_31									
NLCD2001_42									
NLCD2001_41							35.06		
NLCD2001_82		10.62							
AREASQKM		4.66							
LITHO_12		3.78							
NLCD2006_23		2.31							
ROADLEN_KM		1.88							