ARTICLE

Towards a Process Domain-Sensitive Substrate Habitat Model for Sea Lampreys in Michigan Rivers

Thomas M. Neeson,[*][1] **Sara A. Adlerstein, and Michael J. Wiley**

School of Natural Resources and Environment, University of Michigan, 440 Church Street, Ann Arbor, Michigan 48109-1041, USA

Abstract

Habitat mapping is a common and often useful tool in the ecological management of rivers. The complex nature of fluvial processes, however, makes it difficult to predict the reach-scale distribution of substrate habitat from landscape-scale covariates. An option is to identify and partition a data set on boundaries of geomorphic process domains, within which the globally complex relationships between landscape, climate, and instream habitat may potentially be approximated by a simpler model. In this study, we used regression trees as a machine learning method for partitioning and identifying useful strata in a geographically extensive substrate habitat model for larvae of the sea lamprey *Petromyzon marinus*, an invasive and economically harmful species in the Laurentian Great Lakes. We used field survey data from over 5,000 substrate habitat transects collected in 43 watersheds of the Lower Peninsula of Michigan, and we created a geographic database of geographical information systems-derived covariates that represent the principal geomorphic influences on substrate habitat. We created three trees in which tree splits delineated (1) spatially contiguous units, (2) noncontiguous units defined by values of the covariates, and (3) both contiguous and noncontiguous units. The adjusted R^2 values of the three trees were 0.30, 0.30, and 0.32, respectively, and all three trees outperformed a single model fitted to the entire data set and a set of models fitted to each watershed individually. The trees identified useful stratifications of Michigan's Lower Peninsula, important geomorphic influences on substrate habitat, and variation in the influence of geomorphic processes on substrate habitat across our study region. Conservation and management applications of our model predictions and tree-based stratifications include sea lamprey population modeling, habitat survey design, and evaluation of dam removal.

Reach-scale maps of riverine species' preferred habitats play a variety of useful roles in conservation and management (Lunetta et al. 1997; Seelbach et al. 1997; Delong and Brusven 2006; Brenden et al. 2008). For some species, key dimensions of their habitat (e.g., temperature and flow; Zorn et al. 2002; Brenden et al. 2008) are largely predictable from landscape-scale covariates. For these species, it is possible to create statistical models that predict reach-scale habitat from geographic data sets of landscape-scale covariates (Wang et al. 1997; Wehrly et al. 2009) or implicitly model habitat by directly linking landscape characteristics to fish distribution (Steen et al. 2008). Species whose preferred habitats are linked to site-specific substrate characteristics, however, present a unique challenge owing to the complexity of the relationships between landscape, climate, and the geomorphic processes that drive substrate distribution.

Our goal was to use recent developments in fluvial geomorphology as a conceptual basis for predicting reach-scale distributions of preferred substrate habitat for larval sea lampreys *Petromyzon marinus*. The sea lamprey is an invasive and parasitic fish species in the Laurentian Great Lakes and has been linked to the collapse of native lake trout *Salvelinus namaycush* (Smith and Tibbles 1980) and widespread changes in fish community structure (Christie 1974; Smith and Tibbles 1980). As such, sea lampreys are the target of extensive and costly control programs (Applegate 1950; Smith and Tibbles 1980;

Christie et al. 2003). Predictive models of the distribution of larval habitat have potential application in population modeling, in guiding field assessment and lampricide treatment surveys, and particularly in evaluating consequences of dam removal and access for adult sea lampreys to stream reaches upstream from dams. Larval sea lampreys (ammocoetes) prefer habitat that is characterized primarily by soft, burrowable substrate of fine sands (Applegate 1950; Malmqvist 1980; Potter et al. 1986). These substrate preferences were confirmed in laboratory choice experiments (Lee 1989) and thus qualify as a true habitat preference (cf. Rosenfeld 2003). Our analysis focused only on the distribution of preferred substrates (hereafter, "substrate habitat") irrespective of local larval densities and other factors that might at times influence larval distribution, such as oxygen (Potter et al. 1970), temperature (Potter and Beamish 1975), and river network structure (Neeson et al. 2011, 2012).

The distribution of substrate in a reach is controlled by the local energy distributions and sediment budget (Knighton 1998). Efforts to predict reach-scale substrate distribution across large geographic areas must rely on the often tacit assumption that the local-scale energy distributions and sediment budget are correlated with landscape-scale covariates. However, the relationship between reach-scale energy distribution, sediment budget, and landscape-scale covariates is characterized by strong thresholds (Phillips 2003, 2006), nonadditive and higher-order interactions between several ongoing processes (Shreve 1979), and temporal and spatial lags between causes and effects (Madej and Ozaki 1996). Furthermore, substrate habitat models with a large spatial extent may be influenced by variation in geography and physiography (Seelbach et al. 1997), latent climate effects (Langbein and Schumm 1958), and noncausal (spurious) correlations among covariates.

Systems with strong thresholds and complicated interactions often exhibit multiple operating regimes (Eikens and Karim 1999). In geomorphology, this idea is most succinctly stated in the process domain concept (Montgomery 1999). Each process domain is an area or region within which a limited set of geomorphic processes constitutes the dominant influence on instream substrate (Whiting and Bradley 1993; Brardinoni and Hassan 2006). In short, a process domain is the geologic and geomorphic context of a channel unit (Montgomery 1999). Channels identified as belonging to the same process domain should therefore behave similarly with respect to the influence of landscape and climate on instream substrate (Whiting and Bradley 1993). A key consequence is that within each process domain, the globally complex relationships between landscape, climate, and instream substrate habitat may be approximated by a simpler (i.e., more linear) relationship between substrate and some subset of landscape covariates. The ability to identify process domains and partition large data sets accordingly transforms the task of fitting a single, complicated model into the task of fitting a suite of simpler models, each involving fewer covariates. A similar rationale underlies the delineation of hydrologic response units in hydrological modeling (Becker and Pfutzner 1986; Becker and Braun 1999).

The multimodeling framework (Eikens and Karim 1999) provides a useful strategy for constructing a statistical model of a system hypothesized to span multiple process domains. The model-fitting process consists of two steps. First, the data set must be partitioned such that the partitions approximately correspond to the boundaries of the various process domains. Second, a statistical model relating substrate habitat to landscape-scale covariates is fit to the subset of data points in each domain. The resulting model is potentially useful quantitatively as a predictive model and qualitatively as a method for identifying useful stratifications of a large data set or geographic region. Even in cases where the resulting models have low predictive power, the stratifications and key geomorphic variables identified in the process of partitioning the data set may still inform future modeling efforts, sample designs, and identification of units for conservation and management.

Identifying the boundaries of process domains in a large and noisy multivariate data set may be challenging. Previous efforts to identify process domains have used a combination of field work and reasoning from first principles (Whiting and Bradley 1993; Brardinoni and Hassan 2006; Polvi et al. 2011). Neither method for identifying process domains is possible when constructing a spatially extensive model spanning physiographic regions. The class of statistical methods informally known as "machine learning" methods may be useful for discovering the boundaries of process regimes. In particular, regression trees provide a method for completing both steps of the multimodeling procedure outlined by Eikens and Karim (1999): (1) partitioning a data set on the basis of differences in the relationship between substrate habitat and landscape and (2) fitting a regression model to each partition.

In this study, we were interested in exploring the use of regression tree methods to create a process domain predictive model of the distribution of substrate habitat for larval sea lampreys in streams throughout the Lower Peninsula of Michigan. We used a data set that was unique in its spatial extent and geomorphic covariates and so provided a rich setting for testing hypotheses related to the influence of landscape-scale covariates on substrate habitat across process domains. We created regression trees as a machine learning multimodeling method for both partitioning our data set and fitting a linear model to each partition. Our objectives were to (1) identify useful stratifications of the Lower Peninsula of Michigan and (2) understand the key geomorphic processes that influence the distribution of preferred substrate habitat, both across the Lower Peninsula and within identified process domains.

METHODS

The study region was the Lower Peninsula of Michigan (Figure 1). We used field-surveyed substrate habitat data from
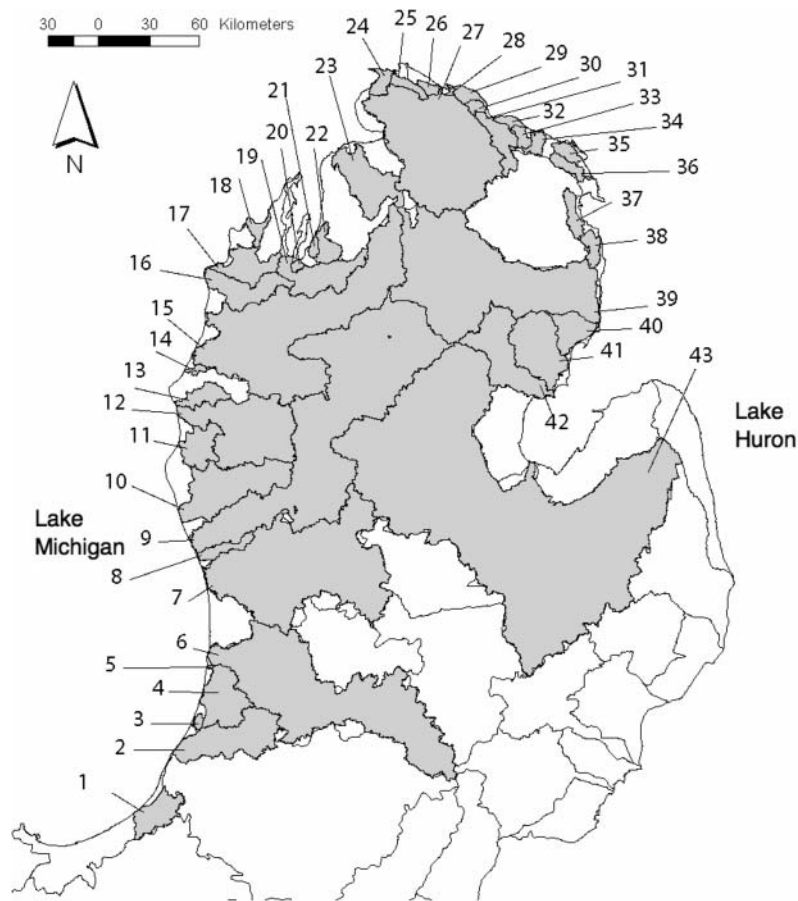
FIGURE 1. The study area, composed of 43 watersheds in the Lower Peninsula of Michigan. Watersheds included in the analysis are shaded, and identification numbers correspond with those in Table 1.

43 watersheds that varied in size from 5 to 16,176 km^2 and exhibited a variety of land use, surficial geology, and hydrology characteristics (Table 1). From these data, we created linear models relating substrate habitat to a set of geographical information systems (GIS)-derived covariates of hydrology, land use, and surficial geology (Table 2). The larval substrate habitat data and geomorphic covariates are described in detail in later subsections.

We created two sets of models by fitting a multiple linear regression to the entire data set and then to each of the 43 tributaries individually. Next, we used regression trees as a method for partitioning a multiple linear regression model by creating three trees with splits based on latitude and longitude, values of the GIS-derived covariates, and both the latitude–longitude and the values of covariates. We compared the explanatory power of all models and examined covariates and strata in the data set identified by the regression tree models.

Modeling Strategy

Our analysis proceeded in five stages. Hereafter, we use the notation "s1–s5" to refer to these five stages.

For the first (null) model (s1), we hypothesized that the relationship between larval sea lamprey substrate habitat and geomorphic covariates may be adequately represented by a single, best-fit linear model across the entire spatial extent and range of covariates of our data set. To test this hypothesis, we fitted one multiple linear regression model relating larval sea lamprey habitat and our GIS-derived covariates to the entire data set.

In the second stage (s2), we hypothesized that the relationship between larval sea lamprey substrate habitat and our covariates varied by watershed. We again assumed that the relationship between substrate habitat and covariates may be represented by linear models, but we also assumed that the relationship between substrate habitat and covariates differs (i.e., has different regression coefficients) among watersheds. To test this set of hypotheses, we fitted a multiple linear regression model to each of the 43 selected watersheds.

For the third-stage models (s3), we again hypothesized that there is spatial variability in the relationship between larval sea lamprey substrate habitat and covariates but that this variability may be best characterized by spatial units other than the watershed. Under this hypothesis, we assumed that the boundaries of process domains are primarily spatial (e.g., between upstream

TABLE 1. Summary statistics for the 43 watersheds in the study area, Michigan's Lower Peninsula (Lake = the Laurentian Great Lake into which each watershed drains [Lake Michigan = 2; Lake Huron = 3]; ID = watershed identification number, corresponding to those in Figure 1; DA = drainage area; n = number of sea lamprey larval habitat data points; type I percentage = mean [with SD in parentheses] percentage of transect width classified as type I [preferred] substrate habitat in quantitative assessment sampling surveys). Mean watershed values for each covariate used as a regressor in our models are also presented (FINALPOWE = stream power; QANN90YLD = 90% exceedance yield or low-flow yield; Dist2Mouth = distance to the river mouth; upDist = distance to the nearest upstream dam or lake; downDist = distance to the nearest downstream dam or lake; FORESTR = percentage of the riparian corridor classified as forested land cover; LACUSTRINE = percentage of the riparian corridor classified as lacustrine surficial geology).

Stream name	Lake	ID	DA (km^2)	n	Type I mean (SD) percentage	FINALPOWE (newtons per second)	QANN90YLD (m^3·s^{-1}·km^{-2})	DIST2MOUTH (m)	upDist (m)	downDist (m)	FORESTR (%)	LACUSTRINE (%)
Galien River	2	1	454.13	182	22.8 (22.5)	18,862	0.00318116	23,747	3,329	6,754	55.5	37.0
St. Joseph	2	2	1,155.61	246	23.7 (20.1)	33,951	0.00466708	62,864	8,823	48,670	71.2	14.5
Brandywine Creek	2	3	41.48	18	30.3 (27.7)	18,075	0.00311000	1,598	4,618	NA	66.4	65.0
Black River	2	4	742.55	30	22.8 (16.9)	14,385	0.00503388	17,978	12,800	7,412	79.0	100.0
Allegan 4 Creek	2	5	5.40	24	19.6 (17.5)	3,061	0.00108700	458	NA	NA	43.5	27.9
Kalamazoo River	2	6	5,039.06	153	11.4 (10.1)	546,366	0.00334988	29,634	1,463	18,059	78.7	92.5
Grand River	2	7	4,944.01	138	16.1 (19.0)	172,140	0.00143214	39,659	15,356	14,630	70.9	62.4
Black Creek	2	8	194.31	36	21.3 (11.4)	12,936	0.00169000	15,219	13,562	10,764	66.6	100.0
Muskegon River	2	9	7,052.16	256	15.6 (15.2)	515,644	0.00497642	45,207	20,114	13,889	69.4	25.2
White River	2	10	1,392.82	80	15.0 (13.8)	86,281	0.00916162	65,668	18,187	7,859	82.0	0.0
Pentwater River	2	11	442.07	42	23.4 (13.4)	11,291	0.00581930	16,448	37,848	NA	79.4	27.2
Pere Marquette River	2	12	1,954.64	443	18.0 (15.4)	75,446	0.00935625	85,136	14,084	28,821	90.8	4.1
Lincoln River	2	13	262.00	46	23.2 (12.6)	16,845	0.00171840	20,435	20,510	18,204	84.6	80.5
Gurney Creek	2	14	48.45	48	15.2 (11.4)	10,600	0.00234124	1,061	6,759	NA	93.0	30.4
Big Manistee River	2	15	5,046.13	389	14.6 (12.6)	291,623	0.00932900	31,216	19,526	21,754	66.6	59.4
Betsie River	2	16	627.38	18	13.3 (13.3)	75,463	0.00368000	12,208	4,914	NA	82.9	34.0
Platte River	2	17	498.38	204	24.5 (17.5)	29,988	0.00453530	8,891	4,202	3,233	71.1	11.7
Good Harbor Creek	2	18	117.83	66	50.7 (32.0)	4,767	0.00644557	4,317	6,695	NA	86.5	81.2
Boardman River	2	19	734.47	34	31.7 (19.5)	76,827	0.00845000	2,604	2,981	NA	32.2	100.0
Mitchell Creek	2	20	40.53	18	37.4 (23.9)	17,554	0.00449201	1,293	3,388	NA	57.5	42.6
Yuba Creek	2	21	58.76	18	19.3 (25.7)	6,536	0.00554350	719	1,279	91	77.1	0.0
Elk Lake Outlet	2	22	217.21	33	11.6 (15.1)	118,609	0.00262760	198	453	NA	14.2	5.8
Jordan River	2	23	860.02	191	18.6 (17.4)	28,005	0.00940700	25,455	11,176	8,563	71.2	71.7
Wycamp Lake Outlet	2	24	138.32	12	21.6 (20.9)	24,624	0.00203778	2,227	3,473	NA	89.2	0.6
Carp Lake River	2	25	119.68	48	16.4 (12.0)	28,449	0.00166000	3,747	11,803	NA	88.3	100.0
Little Black River	3	26	67.93	12	10.5 (15.0)	13,516	0.00180338	1,072	2,004	NA	40.2	80.2
Cheboygan River	3	27	3,867.70	168	12.8 (13.4)	160,882	0.00698000	47,623	31,190	45,141	76.0	46.1
Elliott Creek	3	28	14.46	78	22.8 (12.7)	3,339	0.00231322	3,193	NA	NA	61.2	41.0
Greene Creek	3	29	127.53	60	22.4 (27.2)	2,905	0.00082718	778	NA	NA	84.0	52.8
Black Mallard Creek	3	30	70.87	60	6.8 (9.7)	23,560	0.00253000	4,098	NA	NA	86.9	50.0
Ocqueoc River	3	31	382.65	68	18.8 (14.8)	51,058	0.00465722	22,269	3,030	11,901	84.1	66.1
Schmidt Creek	3	32	81.21	136	21.4 (25.7)	14,879	0.00246032	2,082	NA	NA	78.6	51.5
Trout River	3	33	95.22	166	17.2 (23.2)	35,363	0.00326381	3,298	2,945	1,764	84.2	49.7
Swan River	3	34	113.42	32	5.9 (7.5)	50,176	0.00307678	6,470	NA	5,654	85.1	100.0
Grand Lake Outlet	3	35	124.74	16	72.9 (32.1)	38,809	0.00017776	3,331	328	NA	78.8	100.0
Long Lake Creek	3	36	147.51	31	1.3 (4.6)	61,169	0.00030870	1,488	5,339	NA	73.8	87.7
Devil's River	3	37	185.09	54	11.7 (10.2)	25,089	0.00243093	3,479	3,284	NA	72.5	46.1
Black River	3	38	161.14	96	14.8 (21.6)	20,392	0.00524000	8,895	12,507	NA	74.2	76.3
Au Sable River	3	39	5,197.00	88	8.9 (10.0)	325,189	0.00587000	8,877	10,312	NA	39.1	93.9
Tawas Lake Outlet	3	40	403.51	164	18.4 (14.5)	29,631	0.00232465	9,471	3,437	1,325	73.1	60.4
Au Gres River	3	41	1,012.30	80	16.5 (14.7)	39,543	0.00068721	39,541	12,113	4,695	74.8	48.6
Rifle River	3	42	1,144.29	297	11.7 (11.8)	42,281	0.00350659	73,434	8,255	20,823	72.3	48.3
Saginaw River	3	43	16,176.03	554	9.4 (10.4)	105,446	0.00221350	120,620	10,391	18,501	66.1	65.6

and downstream reaches; Gomi et al. 2002). We created a regression tree as a machine learning method for identifying these units, where partitions in the tree were based on latitude and longitude (hereafter, "spatial tree"). Each terminal node in the regression tree contained a spatially contiguous subset of data points described by a unique multiple linear regression relating

substrate habitat to covariates. If this hypothesis was correct, we expected that the set of regressions based on the regression tree partitions would outperform both models 1 and 2 (i.e., s1 and s2).

In the fourth stage (s4), we hypothesized that the relationship between substrate habitat and covariates is characterized by functional complexity—nonlinearities, thresholds, and

TABLE 2. Summary of substrate habitat and geomorphic variables used in models of preferred substrate habitat for sea lamprey larvae in Michigan's Lower Peninsula. Variable types are response, classification (i.e., used as the basis for regression tree splits), and regressor (used for model fitting). Statistical transformations used for the variables are also presented.

Variable	Description	Variable type	Transformation
prop_type1	Mean proportion of transect length classified as type I (preferred) substrate habitat for larval sea lampreys	Response	$\log_e(\text{prop\_type1} + 0.05)$
YCOORD	Latitude	Classification	None
XCOORD	Longitude	Classification	None
FINALPOWE	Stream power (newtons per second)	Classification, regressor	$\log_e(\text{FINALPOWE})$
QANN90YLD	90% exceedance yield ($m^3 \cdot s^{-1} \cdot km^{-2}$)	Classification, regressor	None
DIST2MOUTH	Distance (m) to the mouth of the river	Classification, regressor	None
FORESTR	Forest cover (percentage of upstream riparian corridor)	Classification, regressor	None
LACUSTRINE	Lacustrine surficial geology (percentage of upstream riparian corridor)	Classification, regressor	None
upDist	Distance (m) to closest upstream dam or lake	Classification, regressor	$\log_e[1/(\text{upDist} + 100)]$
downDist	Distance (m) to closest downstream dam or lake	Classification, regressor	$\log_e[1/(\text{downDist} + 100)]$

interactions between covariates. Under this hypothesis, we assumed that the boundaries of process domains were not necessarily spatial but instead primarily covariate based, analogous to arranging domains on a slope–area plot (Montgomery and Foufoula-Georgiou 1993; Brardinoni and Hassan 2006). One option for modeling such a system is to partition the data set based on values of the covariates such that data points within each partition may be adequately represented by a linear model. To test this hypothesis, we created a regression tree in which partitions were based only on values of the covariates (hereafter, "covariates-only tree").

In a fifth and final modeling stage (s5), which involved the hypothesis that the relationship between substrate habitat and covariates may be characterized by both geographic variability and functional complexity, we created a fifth model that allowed both (1) spatial stratification of the data set and (2) partitions in the data set based on values of the covariates. Under this hypothesis, we assumed that process domains within our study area are at times best identified by their spatial boundaries and at other times best identified by covariate-based boundaries. We created a regression tree as a machine learning method for identifying these partitions, and partitions in the tree were based on latitude and longitude and on values of the covariates (hereafter, "hybrid tree").

Statistical Approach and Analysis

For all of the models, our response variable was the natural logarithm of the mean proportion of transect length that was preferred sea lamprey larval substrate habitat in transects measured by U.S. Fish and Wildlife Service (USFWS) agents (type I habitat in Slade et al. 2003). We used the natural logarithm as our transformation in order to stabilize the variance of the re-

sponse variable with respect to several geomorphic covariates. Covariates were a set of seven GIS-derived variables chosen to represent key geomorphic influences on substrate habitat.

To fit one model (s1) to our entire data set, we selected covariates by using forward and backward stepwise regression. To fit a unique multiple linear regression model to each watershed in our data set (s2), we partitioned the data set by watershed and selected covariates for each watershed's model by using forward and backward stepwise regression. We calculated a single R^2 value for the entire data set by calculating the sum of squares regression for all data points across all watersheds and then dividing the result by the total sum of squares for all data points across all watersheds. We examined the performance of the watershed models in more detail with regression diagnostics on the two best-fitting and two worst-fitting watersheds.

For s3, s4, and s5, we used the GUIDE software package (Loh 2002) to build regression trees in which each terminal tree node (or leaf) contained a multiple linear regression model relating the response (preferred substrate habitat) to some subset of our seven GIS-derived covariates. In GUIDE, the trees were constructed by iteratively splitting the data on values of "classification variables." The classification variable used for each split was the one that minimized the P-value of a chi-square test for curvature in the relationship between substrate habitat and covariates. For each terminal tree node, GUIDE fitted a multiple linear regression model by using forward and backward stepwise regression.

We measured the explanatory power of each tree by calculating the adjusted R^2 value using all points in the data set. We also examined the performance of the trees by performing a detailed diagnostic analysis on the models fitted to the two terminal nodes with the highest adjusted R^2 values (R^2_{adj}) and the two terminal

nodes with the lowest R^2_{adj}. We report the overall importance of each covariate as the number of terminal node regression relationships in which it appeared. We also report the consistency of each covariate—that is, the frequency with which the regression coefficient of the covariate was consistently positive or consistently negative.

Larval Habitat Data

We used an extensive georeferenced database of larval sea lamprey substrate habitat collected during quantitative assessment sampling (QAS) surveys (Slade et al. 2003) conducted by USFWS personnel between 1998 and 2003. During the QAS surveys, substrate was classified visually into three types on the basis of substrate particle size, independent of local larval densities (Slade et al. 2003): type I (preferred) habitat is the substrate most preferred by sea lamprey larvae and is characterized by fine sands and silt; type II (acceptable) habitat is characterized by coarser sands or fine gravel; and type III (unacceptable) habitat is unburrowable (i.e., gravels or coarser substrates). Although this is a subjective classification, there is typically high agreement among observers (Mullett and Bergstedt 2003).

The basic spatial unit in the USFWS database is an "access site." Personnel from USFWS measured substrate habitat along a pair of latitudinal transects at each access site and measured the length of each substrate habitat type along each transect. The location of one of the transects was recorded with a Global Positioning System (GPS) unit. We incorporated 5,027 of these GPS waypoints into our GIS database and corrected the waypoint position when supplementary metadata (e.g., sample date, reach, and access site) contradicted the GPS coordinates. The second transect was located upstream or downstream at a distance of twice the mean stream width.

The two habitat transects associated with each GPS waypoint constituted the basic unit of observation for our study. Our response variable for all modeling analyses was the natural logarithm of the mean proportion of transect length that was preferred larval substrate habitat (i.e., type I habitat) in these two transects.

Geomorphic Covariates

We derived the following geomorphic covariates by using GIS: stream power, low-flow yield, distance to mouth of the river, distance to the nearest upstream lake or dam, distance to the nearest downstream lake or dam, the percentage of the upstream riparian corridor with forested land cover, and the percentage of the upstream riparian corridor with surficial geology characterized as lacustrine deposits. Summary statistics for these covariates by watershed are reported in Table 1.

We chose these specific covariates because they are known to be the dominant influences on substrate particle size (Knighton 1998) and the organization of process domains (Montgomery 1999). Stream power, which is the rate of potential energy expenditure per channel length, determines sediment transport rate (Bagnold 1980) and influences median substrate particle size (Whiting et al. 1999). Stream power is derivable solely from channel slope and drainage area (Whiting et al. 1999; Gorman et al. 2011), two variables that form a basic ordination for process domains (Montgomery and Foufoula-Georgiou 1993; Brardinoni and Hassan 2006) and process-based channel morphologies (Montgomery and Buffington 1997). We hypothesized, therefore, that stream power might be a particularly useful variable for delineating process domains. The 90% exceedance yield (low-flow yield) describes groundwater loading and is therefore a key variable related to the frequency and severity of the flow events that constitute the hydrologic disturbance regime of a channel (Resh et al. 1988). The longitudinal location of a reach, described here by the distance to the river mouth, determines the position of a reach within longitudinally organized process domains (Gomi et al. 2002) and the longitudinal gradient in substrate particle size (Knighton 1980; Brierley and Hickin 1985). Impounded lakes behind dams, and to a lesser degree natural lakes, operate as sediment sinks (Brune 1953). The distance to the nearest upstream lake or dam incorporates the lack of fine sediments below dams (Williams and Wolman 1984; Meade and Parker 1985) into the model, and the distance to the nearest downstream lake or dam incorporates the backwater reaches upstream from the reservoir (Kondolf 1997) into the model. Dams also alter the hydrologic disturbance regime of downstream reaches (Ward and Stanford 1995), an effect that diminishes with distance (Travnichek et al. 1995). Patterns of land use and surficial geology in the upstream catchment determine the overall sediment budget of a channel and thus the amount of fine substrate available to transport processes (summarized by Knighton 1998).

We calculated stream power, distance to the river mouth, distance to the nearest upstream lake or dam, and distance to the nearest downstream lake or dam by using stream path files from the U.S. Geological Survey's National Hydrography Dataset (USGS 2006) updated for greater accuracy for Michigan rivers (Brenden et al. 2006). We obtained modeled low-flow yield, forested land cover, and lacustrine surficial geology from a high-resolution GIS database constructed through a combined effort of the Great Lakes Aquatic Gap Analysis Program (GLSC 2006) and the Classification and Impairment Assessment of Upper Midwest Rivers (Brenden et al. 2006). The database contains 31,817 Michigan stream reaches, and we assigned data to our waypoints by using the data values associated with the stream reach containing each waypoint.

Stream power.–We calculated stream power based on Whiting et al. (1999) as

$$\mathrm{Power} = g \times p \times \mathrm{Flow} \times \mathrm{Slope},$$

where g is the gravitational constant (9.8 m/s^2), p is the density of water, Flow is the 2-year flood flow (m^3/s), and Slope is the stream channel slope (i.e., [vertical distance, m] ÷ [horizontal distance, m]) at the site. We estimated stream channel slope from GIS for an 800-m channel segment enclosing the

waypoint, a channel length previously shown to give good correlation with field-measured water surface slopes (Neeson et al. 2008). Discharge data for estimating the 2-year flood flow were not available for all sites in our analysis, so we estimated 2-year flood flow by using drainage area (Gorman et al. 2011) as follows:

$$\text{Flow} = \text{a} \times \text{Area}^b$$

where $a = 6 \times 10^{-6}$ and $b = 0.8613$; these are regional regression coefficients for northeastern Ohio (Gorman et al. 2011). After preliminary analysis, we selected the natural logarithm of stream power as our covariate.

Low-flow yield.–Low-flow yield is the 90% exceedance flow divided by the upstream drainage area. These data were available in an existing Michigan rivers database, where the 90% exceedance flow was generated from hydrologic models (Brenden et al. 2006).

Distance to the mouth of the river.–Distance to the river mouth was calculated for each of our sample waypoints with two ArcView GIS 3.*x* extensions, Network Analyst (ESRI 1998) and Multiple Closest Facilities, which generated a measurement of the distance along the National Hydrography Dataset stream path shapefile between each sample waypoint and the river mouth.

Distance to the nearest upstream or downstream lake or dam.–We combined a Michigan Department of Natural Resources (MDNR) data set of dams and a MDNR data set of inland lakes to create a single data file of the location of 311 dams and 974 inland lakes in our study region. We used Network Analyst and Multiple Closest Facilities to calculate the distance from each waypoint in our habitat database to the nearest upstream dam or lake and the nearest downstream dam or lake. We transformed both covariates according to the formula:

$$\text{Distance} = \log_e[1/(\text{Distance} + 100)].$$

Land use and geology.–Land use and surficial geology are categorical variables that each exhibits a sum-to-1 constraint (Barringer et al. 1990) and collinearity between categories, so we used only one category of land use and one category of surficial geology as covariates. From an existing data set (Brenden et al. 2006), we chose the percentage of the riparian corridor that was forested (land use) and the percentage of the riparian corridor with lacustrine deposits (surficial geology), the two categories that had the highest correlation with substrate habitat (i.e., our response variable).

RESULTS

Model Performance

The single linear regression fitted to the entire data set (the null model s1) performed poorly ($R^2 = 0.03$). Regression diagnostics indicated that the poor fit was due to high variability in the response variable in the pooled data set that could not be explained by the model covariates. A plot of predicted values against residuals for the entire model exhibited minimal heteroscedasticity, with the exception of artifacts due to the 0, 1 constraint on the response variable (i.e., the mean proportion of preferred substrate habitat). However, more than 99% of the predicted values fell between 0.10 and 0.25, suggesting that while the 0, 1 constraint violated the assumption of constant variance for some points on or near the extremes, it did not inappropriately constrain model fitting or predictions. Plots of residuals against each of the covariates in the model did not indicate that further covariate transformations would improve the fit.

In the s2 analysis, multiple linear regression models fitted to each of the 43 watersheds individually outperformed the single linear regression model and had an overall R^2 value of 0.24. More-detailed regression diagnostics of the two best-fitting and two worst-fitting watersheds showed that the poor fit was not due to correctable patterns in the residuals.

The regression tree created by using latitude and longitude as classification variables (the spatial tree; s3) outperformed models that were fitted to each watershed individually. It had an overall R^2 of 0.30 and contained 23 terminal nodes (Figure 2a). The regression tree created by using the covariates as classification variables (the covariates-only tree; s4) performed similarly ($R^2 = 0.30$) and contained 21 terminal nodes (Figure 3a). The tree created by using both latitude and longitude and the covariates as classification variables (the hybrid tree; s5) performed best (overall $R^2 = 0.32$) and contained 27 terminal nodes (Figure 4a).

Habitat Patterns in Tree Models

All three of the tree models (s3–s5) described similar coarse patterns in the distribution of preferred substrate habitat. For example, all three described a high proportion of preferred substrate habitat in several of the smaller watersheds in the northeastern portion of Michigan's Lower Peninsula (Wycamp Lake Outlet, Elliott Creek, and Greene Creek; watershed identification numbers [IDs] 24, 28, and 29 in Figure 1; mean percentage of preferred substrate habitat = 28.4%). The south branch of the Saginaw River (watershed ID 43 in Figure 1) was predicted to have low proportions of preferred substrate habitat (mean = 2.44%). For other watersheds, the proportion of preferred substrate habitat varied among streams within the watershed. In the Muskegon River (watershed ID 9 in Figure 1), Bigelow Creek (the northeasternmost tributary) had consistently high proportions of preferred substrate habitat (mean = 22.4%), while Brooks Creek (the middle north tributary) had consistently midrange proportions of preferred substrate habitat (mean = 13.6%).

There were also finer-scale differences in the predictions generated by each tree. For example, the covariates-only tree and the hybrid tree predicted a high proportion of preferred substrate habitat in the upstream reaches of the Rifle River (watershed ID 42 in Figure 1), while the spatial tree predicted little or no habitat.
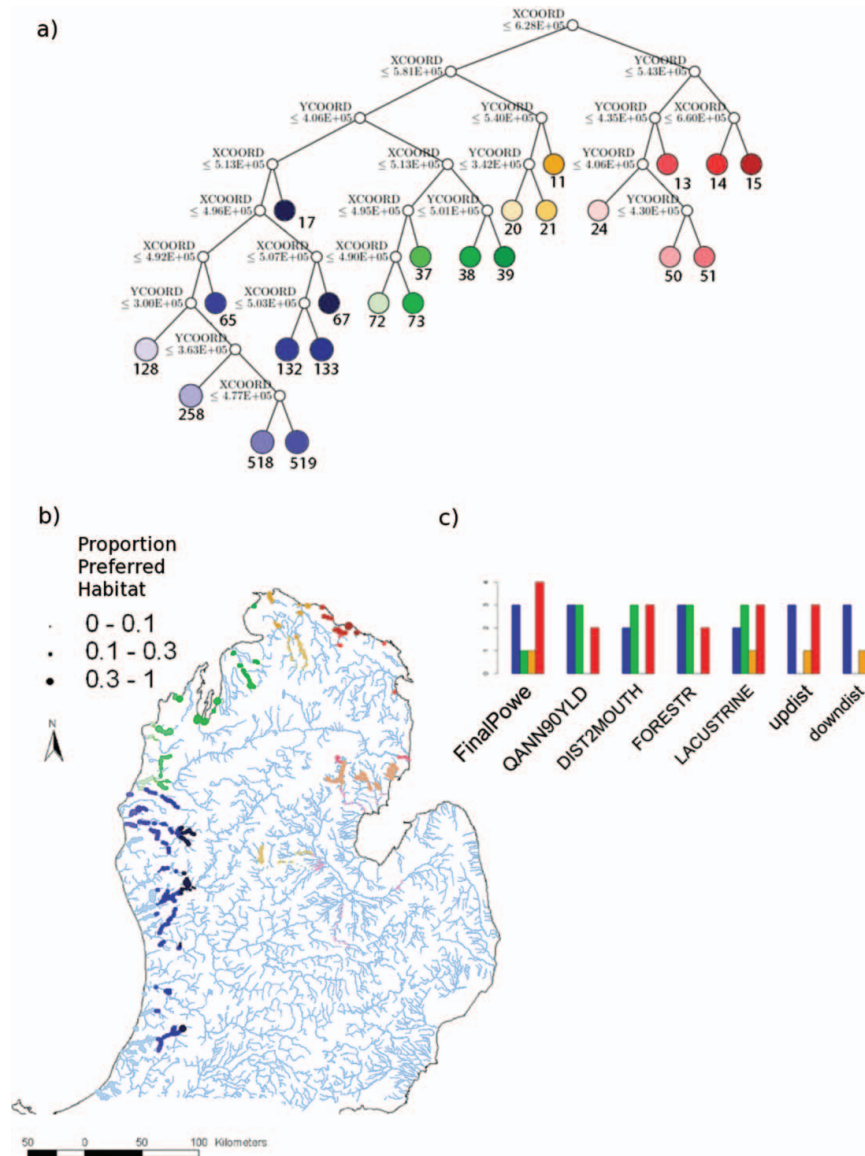
FIGURE 2. **(a)** Fitted regression tree relating preferred substrate habitat for larval sea lampreys to geographical information systems-derived geomorphic covariates, where splits in the tree are based on latitude and longitude (spatial tree, analysis s3; $R^2 = 0.30$). Each tree split is labeled with the value of the covariate that determined the split (YCOORD = latitude; XCOORD = longitude). The node identification number is given below each terminal node and corresponds to individual regression equations. The color of each node corresponds to the colors of waypoints in the **(b)** map of 5,027 waypoints of habitat transect data used to fit the regression tree. The size of each waypoint reflects the predicted proportion of habitat transect length that is preferred larval habitat. Color groups (i.e., blue, green, orange, and red) in the tree and map correspond to the **(c)** bar plot, which shows the number of times that each covariate (defined in Table 2) was used in a terminal node regression relationship for all terminal nodes in that color group. [Figure available online in color.]

Variation in explanatory power within and among the trees is reflected in the R^2 values of the individual nodes. In the spatial tree, the overall R^2 value was 0.30, but the R^2 of individual nodes varied between 0.05 and 0.45. Similarly, the R^2 of individual nodes ranged from 0.04 to 0.51 in the covariates-only tree and from 0.07 to 0.50 in the hybrid tree.

We found no obvious patterns between the R^2 value of a terminal node and the number of data points or the particular covariates selected. More-detailed regression diagnostics of the

two best-fitting and two worst-fitting nodes in the hybrid tree (as measured by R^2_{adj}) showed that the two worst-fitting nodes were not hampered by a need for further transformations or other remedial adjustments. We also saw no obvious relationship between the variance of the response and the fit of the model.

Geomorphic Influences on Substrate Habitat

Covariates differed in their importance (the number of terminal node regression relationships in which they appeared)
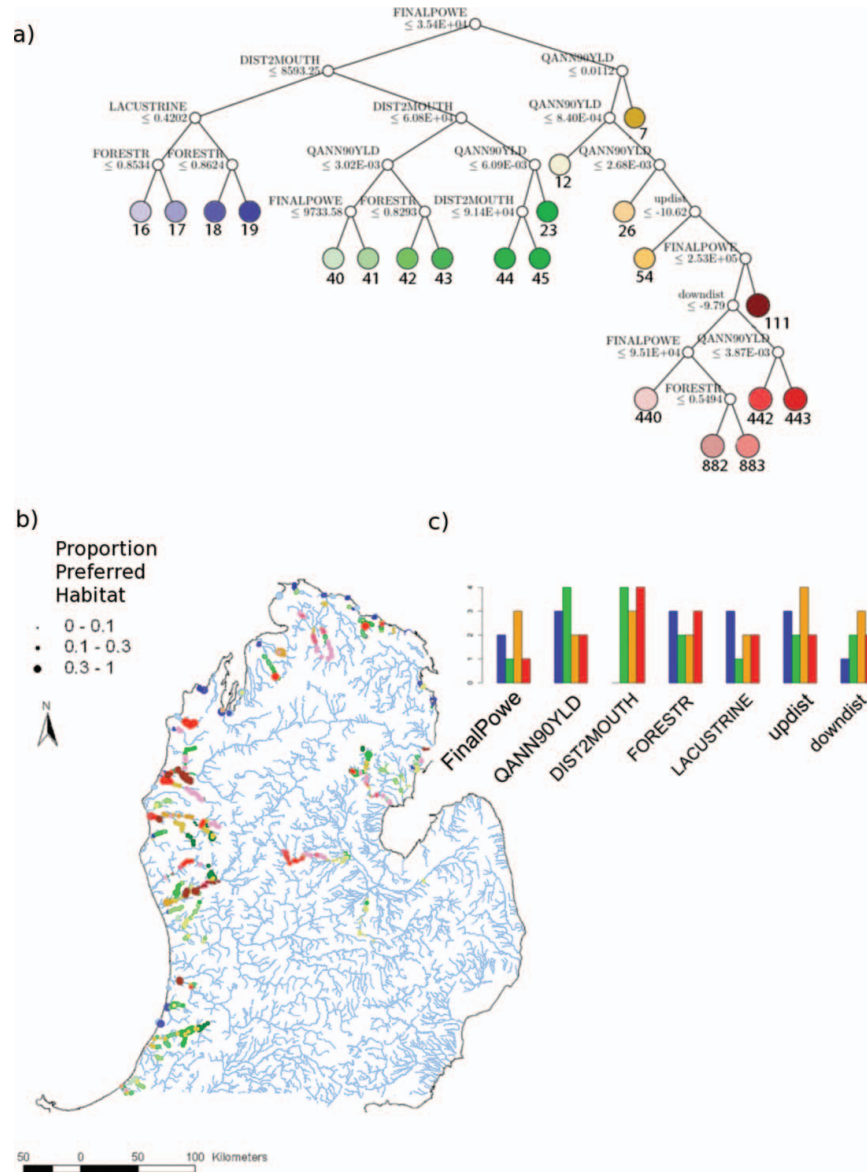
FIGURE 3. Panel descriptions are the same as those in Figure 2, except that splits in the fitted regression tree are based on values of the covariates (covariates-only tree, analysis s4; $R^2 = 0.30$). Covariates are defined in Table 2. [Figure available online in color.]

and their consistency (the frequency at which the coefficient of the covariate was consistently positive or consistently negative; Table 3). Distance to the river mouth and distance to the nearest upstream dam or lake were the most important covariates (occurring in 32 and 31 nodes overall, respectively). Distance to the river mouth was negatively related to preferred substrate habitat in 19 of 32 nodes, while distance to the nearest upstream dam or lake was negatively related to preferred substrate habitat in 16 of 31 nodes. Stream power was the third-most important covariate, occurring in 29 nodes, and had the most consistent relationship with substrate habitat; there was a negative relationship between stream power and the proportion of substrate

habitat in 24 of 29 nodes. The least important covariate, the distance to the nearest downstream dam or lake, occurred in 22 nodes and was positively related to substrate habitat in 12 of 22 cases. For all covariates, their importance and consistency were approximately similar across all three regression trees.

Data Set Partitions

Covariates differed in the frequency with which they were used as the basis for a split within a regression tree (Table 4). Low-flow yield was the most frequently used covariate and served as the basis for eight splits among the three trees, followed by distance to the river mouth (seven splits). Stream power,

TABLE 3. Summary of importance and consistency for covariates (defined in Table 2) in each of the three trees and in total (importance = number of terminal node regression models in which the covariate appeared; consistency = frequency with which the coefficient of the covariate was consistently positive or consistently negative: i.e., number of terminal node regression models in which the covariate had a positive [+] or negative [−] coefficient).

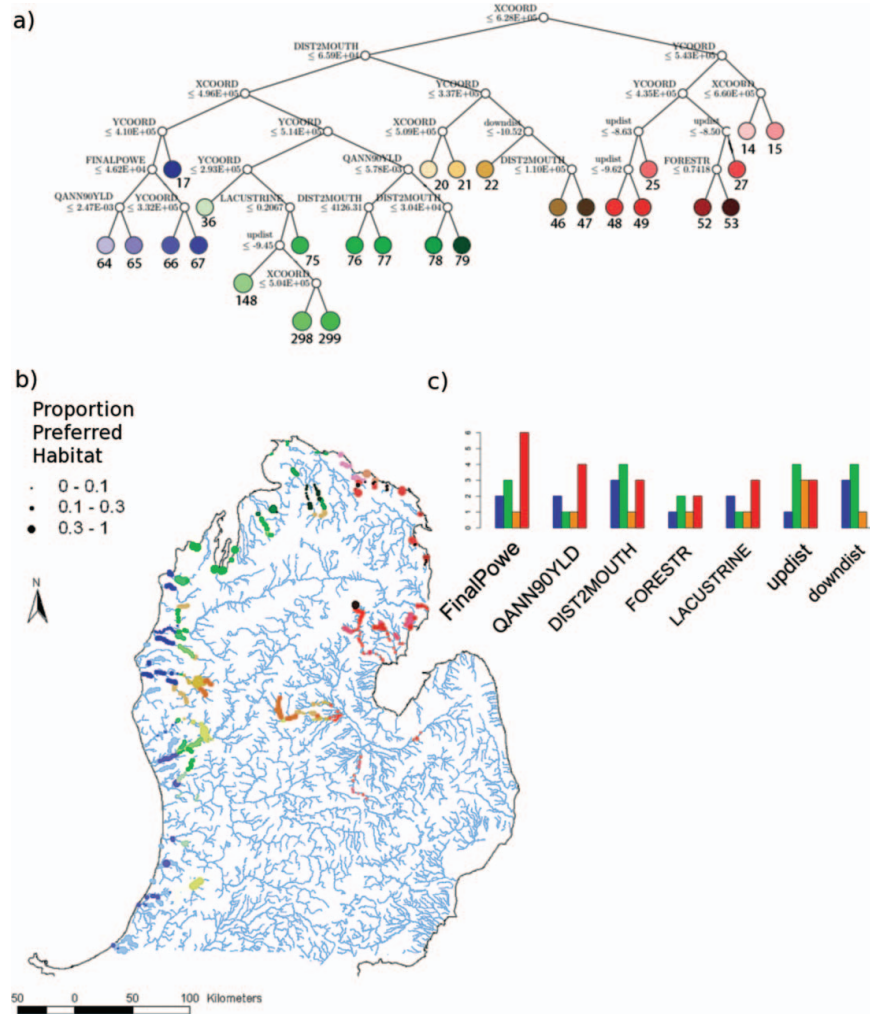| Variable	Spatial tree			Covariates-only tree			Hybrid tree			Total (all trees)		
	Importance	−	+	Importance	−	+	Importance	−	+	Importance	−	+
DIST2MOUTH	9	6	3	11	6	5	12	7	5	32	19	13
upDist	8	5	3	12	5	7	11	6	5	31	16	15
FINALPOWE	9	8	1	8	6	2	12	10	2	29	24	5
QANN90YLD	8	4	4	11	5	6	6	3	3	25	12	13
FORESTR	8	3	5	10	3	7	6	1	5	24	7	17
LACUSTRINE	9	5	4	8	6	2	7	5	2	24	16	8
downDist	5	2	3	9	5	4	8	3	5	22	10	12

FIGURE 4. Panel descriptions are the same as those in Figure 2, except that splits in the fitted regression tree are based on latitude, longitude, and values of the covariates (hybrid tree, analysis s5; $R^2 = 0.32$). Covariates are defined in Table 2. [Figure available online in color.]

TABLE 4. Summary of the number of regression tree partitions based on each classification variable (defined in Table 2) for each of the three trees and in total.

Variable	Spatial tree	Covariates-only tree	Hybrid tree	Total (all trees)
XCOORD	12		5	17
YCOORD	10		6	16
QANN90YLD		6	2	8
DIST2MOUTH		3	4	7
FINALPOWE		4	1	5
FORESTR		4	1	5
upDist		1	4	5
downDist		1	1	2
LACUSTRINE		1	1	2

distance to the nearest upstream dam or lake, and riparian corridor forest were each used for five splits. Low-flow yield, stream power, and riparian corridor forest were more commonly used in the covariates-only tree, while distance to the river mouth and distance to the nearest upstream dam or lake were more commonly used in the hybrid tree. Riparian corridor lacustrine deposits and the distance to the nearest downstream dam or lake were the least frequently used (two splits each).

In the spatial tree, the first tree split divided the data set along a north–south line closely corresponding to a split between the Lake Huron and Lake Michigan drainages. Second-tier splits further divided the western half of the data set along an east–west line and the eastern half of the state along a north–south line. Covariates differed in their importance among these four partitions (Figure 2c). Data points in the northwestern portion of the Lower Peninsula (green points in Figure 2) never included the distance to the nearest upstream or downstream dam or lake as a covariate. Data points in the easternmost portion of the data set (red points in Figure 2) never included distance to the nearest downstream dam or lake as a covariate.

In the covariates-only tree, the topmost split was based on stream power. Second-tier splits were based on distance to the river mouth and low-flow yield. We again saw differences in the importance of covariates between the resulting four subsets of the data set (Figure 3c). Data points with low stream power and near the river mouth (blue points in Figure 3) were never modeled with distance to the river mouth as a covariate and only once used distance to the nearest downstream lake or dam.

The topmost split in the hybrid tree also approximated the split between the Lake Huron and Lake Michigan drainages, and second-tier splits were made on distance to the river mouth and latitude. Data points corresponding to the red terminal nodes in Figure 4a (approximately those points draining to Lake Huron) were never modeled by using distance to the nearest downstream lake or dam (Figure 4c). Data points corresponding to the orange terminal nodes (approximately those points draining to Lake Michigan and having a high distance to the river mouth) were

most frequently modeled by using the distance to the nearest upstream dam or lake.

DISCUSSION

The null linear model fitted to the entire data set (s1) performed poorly. The set of linear models fitted to each watershed individually (s2) had better explanatory power, suggesting that there is spatial variability in the underlying relationship between preferred substrate habitat and geomorphic covariates and that some of this variability occurred by watershed. All three regression trees (s3, s4, and s5) outperformed the watershed-based models, suggesting that tree-identified units are a better basis than watersheds for stratifying the data set. The performance of the covariates-only tree (s4) suggests that partitioning the data set by values of the covariates allows linear models to approximate portions of a more-complex relationship between substrate habitat and covariates. The moderate increase in fit for the hybrid tree (s5) suggests that both spatial and covariate-based partitions on the data set occur simultaneously.

Despite their modest explanatory power (R^2 of about 0.30 for each tree), the regression trees allowed us to (1) identify useful stratifications of the Lower Peninsula of Michigan, (2) identify the most important geomorphic influences on substrate habitat, and (3) describe how the influence of geomorphic covariates on substrate habitat varies across Michigan's Lower Peninsula. The stratifications identified in our analysis might be particularly valuable for the design of stratified field surveys of habitat and other substrate habitat modeling. Our results overall have several specific applications related to the management of sea lampreys in the Great Lakes (see Management Applications below).

Inference from Tree Splits

Stratifications identified by the regression trees represent statistically significant differences in the coefficients of the regression relationships between substrate habitat and covariates. For the spatial tree and the latitude and longitude splits in the hybrid tree, the tree splits delineate spatially contiguous units among which the relationship between substrate habitat and covariates differs. For the covariates-only tree and the covariate-based splits in the hybrid tree, the tree splits represent process-based partitions in the data set within which the complex relationship between substrate habitat and covariates can be better approximated by a linear model.

The partitions identified by our regression tree analysis may inform stratification strategy for future studies and modeling efforts. For example, the topmost split in the spatial and hybrid trees was a split between the eastern and western sides of Michigan, approximately through the center of the Lower Peninsula. This split may be due to latent climate variation in the form of differences in precipitation regime between these two areas (Albert 1995). This hypothesis could be tested with further study, and variation in precipitation could be incorporated into future habitat models explicitly (by introducing a covariate describing precipitation) or implicitly (by stratifying studies or models by

an east–west partition). Second-tier spatial partitions divided the western half of the Lower Peninsula into north and south units and divided the eastern half of the Lower Peninsula into east and west units.

The size of our data set and the number of tree nodes limited our ability to understand the basis of covariate-based tree splits. For example, in a regression tree analysis of habitat for the desert tortoise *Gopherus agassizii*, Andersen et al. (2000) were able to use tree splits to identify interactions between covariates, but that study involved a smaller data set and a smaller geographic area. However, covariates that served as key classification variables in our data set can still be useful for stratification in future studies and models. The top-level splits in the covariates-only tree (stream power, distance to the river mouth, and 90% exceedance yield) and the hybrid tree (distance to the river mouth) probably represent the most important partitioning of the data set for management or future modeling efforts, while partitions farther down the tree could be included if financially feasible.

Relationship between Habitat and Covariates

The regression coefficients of some covariates were consistent in sign across tree nodes and reflected previously known relationships between geomorphology and substrate habitat. Preferred habitat of sea lamprey ammocoetes consists of relatively easily transported sands, so we would expect a negative correlation between the amount of preferred substrate habitat and the sediment transport rate. Stream power had a negative regression coefficient in 24 of 29 terminal nodes, which corresponds to empirical studies demonstrating positive correlations between stream power and sediment transport rate (Bagnold 1980) or median substrate particle size (Whiting et al. 1999). The amount of forest in the riparian corridor had a positive regression coefficient in 17 of 24 terminal nodes, which corresponds to empirical studies describing the negative correlation between forested landscapes and sediment transport rates (Milliman et al. 1987).

Rivers in glaciated landscapes such as Michigan flow through regions of varying land use and geology, and the relative importance of geomorphic drivers of substrate habitat may differ among these regions. Differences in the importance of covariates between strata reflect this variability in the influence of geomorphology on substrate habitat. For example, in the spatial tree, substrate habitat in the northwestern quarter of the Lower Peninsula was never influenced by the distance to the nearest upstream or downstream dam or lake (green points in Figure 2), yet this was an important geomorphic influence on substrate habitat for southwestern Michigan (red points in Figure 2). These patterns of variability in the importance of covariates may be characterized at any of several spatial scales (i.e., tree depths) and may be combined with expert knowledge of regional hydrology to inform future studies and models of habitat.

For other covariates, such as low-flow yield and the distance to the nearest upstream dam or lake, the number of nodes with positive regression coefficients and the number of nodes with negative regression coefficients were approximately equal. This may reflect changes in the relationship between geomorphology and substrate habitat across Michigan's glaciated landscapes. Alternatively, the sign of the relationships may be due to a mismatch between the scales at which we measured covariates and the scales at which they influence substrate habitat. For example, a dam may limit downstream sediment supply (Williams and Wolman 1984; Meade and Parker 1985), but this effect diminishes with distance and the supply of sediment from additional tributaries downstream. When the measured distance from a waypoint to the nearest dam is greater than several kilometers, the covariate may be serving as a general description of geographic position. Distance to the river mouth was chosen to incorporate the gradient of increasing substrate particle size from the river mouth to the river headwaters (Knighton 1980; Brierley and Hickin 1985), but this gradient is rarely smooth, and local convex-upwards shapes and scatter may explain the 13 positive regression coefficients in 32 terminal nodes.

Management Applications

A general management objective for sea lamprey control in the Great Lakes is to increase the number of ammocoetes killed per dollar spent (a cost : kill ratio; Christie et al. 2003; Slade et al. 2003), and proposed modifications to the sampling protocol used to rank streams for lampricide treatment are typically evaluated in terms of their cost : kill impact. Until 2008, estimates of both larval density and total amount of larval habitat were necessary for estimating the expected production of parasitic-phase sea lampreys in each reach; therefore, improving the estimates of either would be beneficial for the accuracy of stream rankings. However, simply increasing the amount of field sampling performed is not a feasible option. Regarding larval density estimates, Hansen et al. (2003) calculated that field sampling efforts ". . . would need to be increased 52% to increase precision of the larval density to 50% of the mean, and by more than 240% to increase precision of the larval density estimate to 33% of the mean." The calculations would be different for habitat, but the calculations for larvae as presented by Hansen et al. (2003) demonstrate the general point that large increases in field sampling efforts are needed to increase the precision of measurements.

Although the QAS protocol has recently been replaced by a rapid assessment protocol that uses historical data in place of additional field sampling of substrate habitat (Hansen and Jones 2008), accurate estimates of the distribution and amount of preferred ammocoete habitat are still desirable. The switch from the QAS protocol to the rapid assessment protocol was made on the basis of a cost–benefit analysis, but good predictive substrate habitat models that complement and guide field sampling efforts could significantly change the costs associated with measuring habitat.

Predictive substrate habitat models might be more useful for predicting the distribution and amount of habitat upstream of dams that are being considered for removal. Adult sea lampreys are constrained below most dams during their spawning runs,

and dam removal could open large areas to infestation. Dam removal is being considered for many rivers to improve ecological integrity (Lavis et al. 2003) and is a pressing management concern given the more than 2,000 existing dams in the Lower Peninsula of Michigan (MDNR, unpublished data). Predictive substrate habitat models that complement field surveys provide a cost-effective way to estimate the amount of preferred ammocoete habitat that will be opened to infestation.

Given the modest predictive power of our models (R^2 of approximately 0.30 for all trees), they may best be used as the basis for a stratified sampling regime (Guisan and Zimmerman 2000; Cole et al. 2001) that guides, rather than replaces, existing field sampling efforts. Specifically, the trees identify domains where the explanatory power of our models is relatively good ($R^2 > 0.50$) as well as domains where the explanatory power of our models is low ($R^2 < 0.10$). These two extremes will require different allocation of field sampling efforts to arrive at the same degree of confidence in estimates of larval habitat within a reach. Tree-identified domains may therefore serve as the basis for efficiently allocating field survey efforts.

Similarly, tree-identified domains may be used to provide a geomorphic context for other process-based substrate habitat models, such as those based on local stream geometry (Neeson et al. 2007) or a combination of channel slope and basin size (Gorman et al. 2011; Wilkins and Snyder 2011). A model-averaging approach may then generate improved predictions or field sampling strategies.

Despite differences in their predictive power, each step in the analysis contributed to a better understanding of the processes and spatial units influencing substrate habitat. The spatial and hybrid trees identified geographic areas that could serve as the basis for management units or stratifications for partitioning future habitat models. The presence or absence of covariates in the regression relationship that described each of these nodes highlighted the critical influences on larval substrate habitat within that management unit. In addition, the trees identified key geomorphic influences on habitat and key stratifying geomorphic variables, both of which may be used to understand the processes influencing habitat and to inform future modeling efforts.

ACKNOWLEDGMENTS

REFERENCES

Albert, D. A. 1995. Regional landscape ecosystems of Michigan, Minnesota, and Wisconsin: a working map and classification. U.S. Department of Agriculture Forest Service, General Technical Report NC-178, North Central Forest Experiment Station, St. Paul, Minnesota.

Andersen, M. C., J. M. Watts, J. E. Freilich, S. R. Yool, G. W. Wakefield, J. F. McCauley, and P. B. Fahnestock. 2000. Regression tree modeling of desert tortoise habitat in the central Mojave Desert. Ecological Applications 10:890–900.

Applegate, V. C. 1950. Natural history of the sea lamprey (*Petromyzon marinus*) in Michigan. U.S. Fish and Wildlife Service Special Scientific Report Fisheries 55.

Bagnold, R. A. 1980. An empirical correlation of bedload transport rates in flumes and natural rivers. Proceedings of the Royal Society of London 372A:453–473.

Barringer, T. H., D. Dunn, W. A. Battaglin, and E. F. Vowinkel. 1990. Problems and methods involved in relating land use to ground-water quality. Water Resources Bulletin 26:1–9.

Becker, A., and P. Braun. 1999. Disaggregation, aggregation, and spatial scaling in hydrological modeling. Journal of Hydrology 217:239–252.

Becker, A., and B. Pfutzner. 1986. Identification and modeling of river flow reductions caused by evapotranspiration losses from shallow groundwater areas. IAHS Publication 156.

Brardinoni, F., and M. A. Hassan. 2006. Glacial erosion, evolution of river long profiles, and the organization of process domains in mountain drainage basins of coastal British Columbia. Journal of Geophysical Research 111:F01013.

Brenden, T. O., R. D. Clark, A. R. Cooper, P. W. Seelbach, L. Wang, S. Aichele, E. G. Bissell, and J. S. Stewart. 2006. A GIS framework for collecting, managing, and analyzing multiscale landscape variables across large regions for river conservation and management. Pages 49–74 *in* R. M. Hughes, L. Wang, and P. W. Seelbach, editors. Landscape influences on stream habitat and biological assemblages. American Fisheries Society, Symposium 48, Bethesda, Maryland.

Brenden, T. O., L. Wang, and P. W. Seelbach. 2008. A river valley segment classification of Michigan streams based on fish and physical attributes. Transactions of the American Fisheries Society 137:1621–1636.

Brierley, G. J., and E. J. Hickin. 1985. The downstream gradation of particle sizes in the Squamish River, British Columbia. Earth Surface Processes and Landforms 10:597–606.

Brune, G. M. 1953. The trap efficiency of reservoirs. Transactions, American Geophysical Union 34:407–418.

Christie, G. C., J. V. Adams, T. B. Steeves, J. W. Slade, D. W. Cuddy, M. F. Fodale, R. J. Young, M. Kuc, and M. L. Jones. 2003. Selecting Great Lakes streams for lampricide treatment based on larval sea lamprey surveys. Journal of Great Lakes Research 29(Supplement 1):152–160.

Christie, W. J. 1974. Changes in the fish species composition of the Great Lakes. Journal of the Fisheries Research Board of Canada 31:827–854.

Cole, R. G., T. R. Healy, M. L. Wood, and D. M. Foster. 2001. Statistical analysis of spatial pattern: a comparison of grid and hierarchical sampling approaches. Environmental Monitoring and Assessment 69:85–99.

Delong, M. D., and M. A. Brusven. 2006. Classification and spatial mapping of riparian habitat with applications toward management of streams impacted by nonpoint source pollution. Environmental Management 15:565–571.

Eikens, B., and M. N. Karim. 1999. Process identification with multiple neural network models. International Journal of Control 72:576–590.

ESRI (Environmental Systems Research Institute). 1998. ArcView network analyst: an ESRI white paper. ESRI, Redlands, California.

Gomi, T., R. C. Sidle, and J. S. Richardson. 2002. Understanding processes and downstream linkages of headwater systems. BioScience 52:905–916.

Gorman, A., P. J. Whiting, T. M. Neeson, and J. F. Koonce. 2011. Channel substrate prediction from GIS for habitat estimation in Lake Erie tributaries. Journal of Great Lakes Research 37:725–731.

Guisan, A., and N. E. Zimmerman. 2000. Predictive habitat distribution models in ecology. Ecological Modelling 135:147–186.

Hansen, G. J. A., and M. L. Jones. 2008. A rapid assessment approach to prioritizing streams for controls of Great Lakes sea lamprey (*Petromyzon marinus*): a case study in adaptive management. Canadian Journal of Fisheries and Aquatic Sciences 65:2471–2484.

Hansen, M. J., J. V. Adams, D. W. Cuddy, J. M. Richards, M. F. Fodale, G. L. Larson, D. J. Ollila, J. W. Slade, T. B. Steeves, R. J. Young, and A. Zerrenner. 2003. Optimizing larval assessment to support sea lamprey control in the Great Lakes. Journal of Great Lakes Research 29(Supplement 1):766–782.

Knighton, A. D. 1980. Longitudinal changes in size and sorting of streambed material in four English rivers. Geological Society of America Bulletin 91:55–62.

Knighton, D. 1998. Fluvial forms and processes: a new perspective. Wiley, New York.

Kondolf, G. M. 1997. Hungry water: effects of dams and gravel mining on river channels. Environmental Management 21:533–551.

Langbein, W. B., and S. A. Schumm. 1958. Yield of sediment in relation to mean annual precipitation. Transactions, American Geophysical Union 39:1076–1084.

Lavis, D. S., A. Hallett, E. M. Koon, and T. C. McAuley. 2003. History of and advances in barriers as an alternative method to suppress sea lampreys in the Great Lakes. Journal of Great Lakes Research 29(Supplement 1): 362–372.

Lee, D. S. 1989. Quantified laboratory assessment of larval lamprey substrate habitat selection. Great Lakes Fishery Commission, Research Completion Report, Ann Arbor, Michigan.

Loh, W. Y. 2002. Regression trees with unbiased variable selection and interaction detection. Statistica Sinica 12:361–386.

Lunetta, R. S., B. L. Cosentino, D. R. Montgomery, E. M. Beamer, and T. J. Beechie. 1997. GIS-based evaluation of salmon habitat in the Pacific Northwest. Photogrammetric Engineering and Remote Sensing 63:1219–1229.

Madej, M. A., and V. Ozaki. 1996. Channel response to sediment wave propagation and movement, Redwood Creek, California, USA. Earth Surface Processes and Landforms 21:911–927.

Malmqvist, B. 1980. Habitat selection of larval brook lampreys (*Lampetra planeri,* Bloch) in a south Swedish stream. Oecologia 45:35–38.

Meade, R. H., and R. S. Parker. 1985. Sediment in rivers in the United States. U.S. Geological Survey Water-Supply Paper 2275:49–60.

Milliman, J. D., Y. S. Qin, M. E. Ren, and Y. Saito. 1987. Man's influence on the erosion and transport of sediment by Asian rivers: the Yellow River (Huanghe) example. Journal of Geology 95:751–762.

Montgomery, D. R. 1999. Process domains and the river continuum. Journal of the American Water Resources Association 35:397–410.

Montgomery, D. R., and J. M. Buffington. 1997. Channel-reach morphology in mountain drainage basins. Geological Society of America Bulletin 109:596–611.

Montgomery, D. R., and E. Foufoula-Georgiou. 1993. Channel network source representation using digital elevation models. Water Resources Research 29:3925–3934.

Mullett, K. M., and R. A. Bergstedt. 2003. Agreement among observers classifying larval sea lamprey (*Petromyzon marinus*) habitat. Journal of Great Lakes Research 29(Supplement 1):183–189.

Neeson, T. M., A. Gorman, P. J. Whiting, and J. F. Koonce. 2008. Factors affecting accuracy of stream channel slope estimates derived from geographic information systems. North American Journal of Fisheries Management 28:722–732.

Neeson, T. M., J. F. Koonce, and P. J. Whiting. 2007. Predicting sea lamprey (*Petromyzon marinus*) habitat using geographic information systems. Journal of Great Lakes Research 33:546–553.

Neeson, T. M., M. J. Wiley, S. A. Adlerstein, and R. L. Riolo. 2011. River network structure shapes interannual feedbacks between adult sea lamprey migration and larval habitation. Ecological Modelling 222:3181–3192.

Neeson, T. M., M. J. Wiley, S. A. Adlerstein, and R. L. Riolo. 2012. How river network structure and habitat availability shape the spatial dynamics of larval sea lampreys. Ecological Modelling 226:62–70.

Phillips, J. D. 2003. Sources of nonlinearity and complexity in geomorphic systems. Progress in Physical Geography 27:1–23.

Phillips, J. D. 2006. Evolutionary geomorphology: thresholds and nonlinearity in landform response to environmental change. Hydrology and Earth System Sciences 10:731–742.

Polvi, L. E., E. E. Wohl, and D. M. Merritt. 2011. Geomorphic and process domain controls on riparian zones in the Colorado front range. Geomorphology 125:504–516.

Potter, I. C., and F. W. H. Beamish. 1975. Lethal temperatures in ammocoetes of four species of lampreys. Acta Zoologica 56:85–91.

Potter, I. C., B. J. Hill, and S. Gentleman. 1970. Survival and behavior of ammocoetes at low oxygen tensions. Journal of Experimental Biology 53:59–73.

Potter, I. C., R. W. Hilliard, J. S. Bradley, and R. J. McKay. 1986. The influence of environmental variables on the density of larval lampreys in different seasons. Oecologia 70:433–440.

Resh, V. H., A. V. Brown, A. P. Covich, M. E. Gurtz, H. W. Li, G. W. Minshall, S. R. Reice, A. L. Sheldon, J. B. Wallace, and R. C. Wissmar. 1988. The role of disturbance in stream ecology. Journal of the North American Benthological Society 7:433–455.

Rosenfeld, J. 2003. Assessing the habitat requirements of stream fishes: an overview and evaluation of different approaches. Transactions of the American Fisheries Society 132:953–968.

Seelbach, P. W., M. J. Wiley, J. C. Kotanchik, and M. E. Baker. 1997. A landscape-based ecological classification system for river valley segments in lower Michigan (MI-VSEC version 1.0). Michigan Department of Natural Resources, Fisheries Division, Fisheries Research Report 2036, Lansing.

Shreve, R. L. 1979. Models for prediction in fluvial geomorphology. Mathematical Geology 11:165–173.

Slade, J. W., J. V. Adams, G. C. Christie, D. W. Cuddy, M. F. Fodale, J. W. Heinrich, H. R. Quinlan, J. G. Weise, J. W. Weisser, and R. J. Young. 2003. Techniques and methods for estimating abundance of larval and metamorphosed sea lampreys in Great Lakes tributaries, 1995 to 2002. Journal of Great Lakes Research 29(Supplement 1):137–151.

Smith, B. R., and J. J. Tibbles. 1980. Sea lamprey (*Petromyzon marinus*) in Lakes Huron, Michigan, and Superior: history of invasion and control, 1936–1978. Canadian Journal of Fisheries and Aquatic Sciences 37:1780–1801.

Steen, P. J., T. G. Zorn, P. W. Seelbach, and J. S. Schaeffer. 2008. Classification tree models for predicting distributions of Michigan stream fish from landscape variables. Transactions of the American Fisheries Society 137:976–996.

Travnichek, V. H., M. B. Bain, and M. J. Maceina. 1995. Recovery of a warmwater fish assemblage after the initiation of a minimum-flow release downstream of a hydroelectric dam. Transactions of the American Fisheries Society 124:836–844.

USGS (U.S. Geological Survey). 2006. National hydrography dataset. USGS, Washington, D.C. Available: nhd.usgs.gov. (March 2011).

Wang, L., J. Lyons, P. Kanehl, and R. Gatti. 1997. Influences of watershed land use on habitat quality and biotic integrity in Wisconsin streams. Fisheries 22(6):6–12.

Ward, J. V., and J. A. Stanford. 1995. Ecological connectivity in alluvial river ecosystems and its disruption by flow regulation. Regulated Rivers: Research and Management 11:105–119.

Wehrly, K. E., T. O. Brenden, and L. Wang. 2009. A comparison of statistical approaches for predicting stream temperatures across heterogeneous landscapes. Journal of the American Water Resources Association 45:986–997.

Whiting, P. J., and J. B. Bradley. 1993. A process-based classification system for headwater streams. Earth Surface Processes and Landforms 18:603–612.

Whiting, P. J., J. F. Stamm, D. B. Moog, and R. L. Orndorff. 1999. Sediment-transporting flows in headwater streams. Geological Society of America Bulletin 111:450–466.

Wilkins, B. C., and N. P. Snyder. 2011. Geomorphic comparison of two Atlantic coastal rivers: toward an understanding of physical controls on Atlantic salmon habitat. River Research and Applications 27:135–156.

Williams, G. P., and M. G. Wolman. 1984. Downstream effects of dams on alluvial rivers. U.S. Geological Survey Professional Paper 1286.

Zorn, T. G., P. W. Seelbach, and M. J. Wiley. 2002. Distributions of stream fishes and their relationship to stream size and hydrology in Michigan's Lower Peninsula. Transactions of the American Fisheries Society 131:70–85.