

Patch-based Models to Predict Species Occurrence: Lessons from Salmonid Fishes in Streams

Jason B. Dunham, Bruce E. Rieman, and James T. Peterson

Environmental heterogeneity often produces patchy or discontinuous distributions of organisms. Even broadly distributed species show localized peaks of abundance (Maurer 1999). This is particularly obvious in stream ecosystems, where patch dynamics is a dominant theme (Pringle et al. 1988). Features of the environment that may influence species occurrence in streams are believed to result from a hierarchy of physical processes operating within drainage basins. This idea has formed the basis of several classification schemes for stream habitats (e.g., Frissell et al. 1986; Hawkins et al. 1993; Imhof et al. 1996; Naiman 1998; see Morrison and Hall (Chapter 2) for definition of “habitat”). These classifications provide a useful framework for understanding physical processes that generate stream habitat over areas of varying size and spatial resolution, but they do not explicitly consider how individual species actually perceive or utilize these patchy environments.

To be most useful, patches should be clearly defined by associations between a biological response (e.g., reproduction, migration, feeding) and environmental variability (Addicott et al. 1987; Kotliar and Wiens 1990). Classifications of aquatic habitat based purely on physical characteristics or subdivision of watersheds into arbitrary segments may not adequately describe “patchiness” from an organism’s point of view. Lack of attention to realistic scaling of environmental

variation and biological responses can produce weak or misleading inferences (Goodwin and Fahrig 1998). Our focus in this chapter is on definition of patches suitable for supporting local breeding populations. This is a key prerequisite for applying ideas from metapopulation and landscape ecology to predicting species occurrence.

Here, we review our attempts to develop patch-based classifications of aquatic habitat and models to predict occurrence of salmonid fishes in streams. We begin with a brief overview of the concept of patchiness. Next, we outline criteria to define the biological response of interest: occurrence of local populations. We then describe models to predict the distribution of local populations within stream basins. These models allow delineation of patches of suitable habitat within watersheds and definition of patch structuring. Patterns of patch structuring and characteristics of individual patches provide the basis for modeling occurrence of local populations. We compare patch-based models of occurrence for two threatened salmonids: bull trout (*Salvelinus confluentus*) and Lahontan cutthroat trout (*Oncorhynchus clarki henshawi*). Finally, we compare our results to alternative approaches to predict occurrence of salmonids and discuss implications of a patch-based approach that should be generally relevant for developing models of species occurrence.

The Concept of Patchiness

The term “patch” has been applied in numerous contexts in ecology (e.g., Pickett and White 1985; McCoy and Bell 1991; Pickett and Rogers 1997; Morrison and Hall, Chapter 2). Our definition of patches parallels the concept of ecological neighborhoods introduced by Addicott et al. (1987). Ecological neighborhoods are defined by a specific biological response and not by an arbitrary temporal or spatial scale or by a perceived boundary or control imposed on the system. A “patch” corresponds to limits or boundaries of environmental conditions that can support a biological response. Patches of environmental conditions potentially suitable to support local populations of a species are often the focus in landscape and metapopulation ecology.

Kotliar and Wiens (1990) provided a general framework for defining patch structure. Given that a biological response is observed to occur within a definable spatial frame, patch structure can be characterized by (1) the degree to which patches can be distinguished from each other and the surrounding environment (patch contrast), and (2) how patches are spatially aggregated. Patch structuring may be characterized by a nested or hierarchical pattern and may vary widely among biological responses.

Patch structuring is not directly synonymous with a specific temporal or spatial scale. For example, patches defined here may vary by an order of magnitude or more in size (patch area). It is definition of common biological responses and environmental criteria for determining patch structure, not spatial or temporal scale per se, that provides a foundation for patch-based models of species occurrence that may be generalized within and among species.

Defining a Biological Response

We were interested in predicting occurrence of fish in patches of habitat suitable for local breeding populations. Patches suitable for local populations should correspond to locations where population growth can be attributed primarily to in situ reproduction, rather than immigration (Addicott et al. 1987). Limited demographic interaction among local populations im-

plies some degree of reproductive (genetic) isolation. Spatial isolation of spawning and rearing habitat for salmonids is reinforced by strong natal homing (Quinn 1993), and patches of suitable habitat may therefore support relatively discrete local populations. Ultimately, it would be desirable to use multiple sources of information to delineate local breeding populations. Several studies have demonstrated the limitations of using limited genetic or demographic information alone to infer population structuring (Ims and Yoccoz 1997; see Utter et al. 1992, 1993 for salmonid examples).

Unfortunately, detailed genetic and demographic data are not available for most systems. For salmonids, we have defined patches of suitable habitat by modeling the distribution limits of smaller, presumably “pre migratory” or resident individuals within streams. Larger juvenile and adult salmonids may adopt migratory life histories (Northcote 1997, Fig. 26.1) and range far outside of spawning and rearing areas, but their existence ultimately depends on spawning and rearing habitat. Our delineation of patches for salmonids, then, is based principally on ecological information and an assumption of natal homing. Population genetic analysis (e.g., Kanda

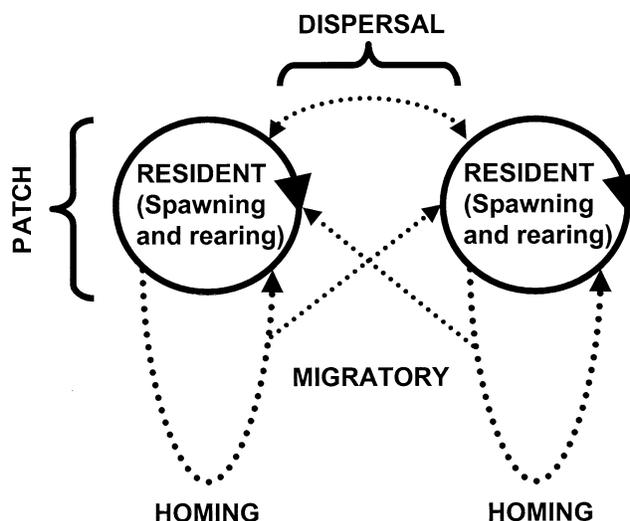


Figure 26.1. Simplified schematic of life-history variation in salmonid fishes. Fish with a “resident” life history spend their entire lives within spawning and rearing areas. Migratory fish use habitats outside of spawning and rearing areas but return faithfully (homing) to breed in natal areas. Some dispersal is possible among both life-history types. Our definition of patches corresponds to the extent of spawning and rearing.

1998; Spruell et al. 1999) for some systems indicates genetic divergence does correspond to juvenile distributions. Our approach to defining patches appears to be a reasonable approximation, but detailed demographic and/or genetic data will be necessary to confirm the structure of any system (Haila et al. 1993; Rieman and Dunham 2000).

Models of Distribution Limits and Patch Delineation

Unlike terrestrial habitats, streams are generally viewed as one-dimensional systems in terms of fish distributions and dispersal. Therefore, boundaries of habitat patches may be delineated in an up- and/or downstream direction. Many factors can potentially limit the distribution of spawning and rearing habitat for salmonids, including natural and artificial dispersal barriers, water temperature, interactions with non-native salmonids and other fishes, human disturbance, and geomorphic influences. These factors are often not independent. For example, interspecific interactions mediated by water temperature may influence longitudinal distributions of species within streams (De Staso and Rahel 1994; Taniguchi et al. 1998).

In the case of bull trout and cutthroat trout, spawning and early rearing usually occur in upstream or headwater habitats (often fourth-order streams or smaller), so we were particularly interested in factors that determine downstream distribution limits of juveniles. Our two study areas are located at the southern margin of the range for both species, where unsuitably warm summer water temperatures in streams are probably an important factor limiting the amount of suitable habitat (Rieman et al. 1997; J. B. Dunham and B. E. Rieman unpublished data). Local populations of these species in other areas may be delineated by different habitat characteristics, such as availability of high-quality spawning habitat (Baxter et al. 1999; also see Geist and Dauble 1998), barriers (e.g., dams, waterfalls, subsurface flow), and sharp transitions in habitat that occur as tributary streams flow into larger streams or lakes.

There are many recent examples of attempts to classify aquatic habitat for salmonids based on different indicators related to variability in stream tempera-

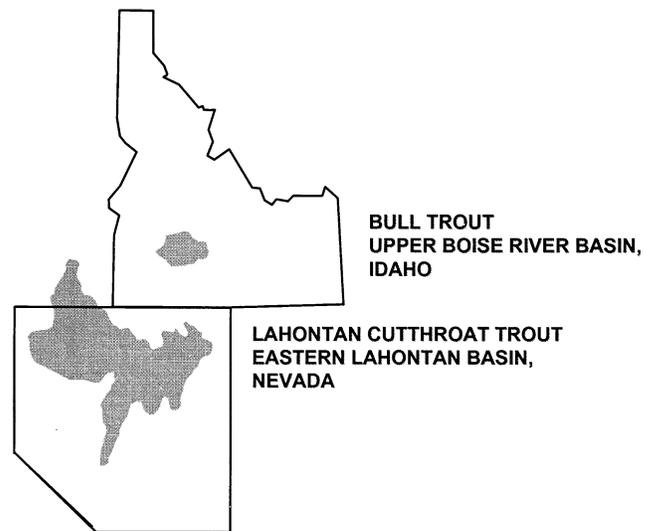


Figure 26.2. Map of study areas: the upper Boise River Basin, Idaho, and the eastern Lahontan Basin, Nevada.

ture. Various researchers have classified thermally suitable habitat from variation in groundwater (Meisner 1990; Nakano et al. 1996), air (Keleher and Rahel 1996), and surface water temperatures (Eaton et al. 1995; Rahel et al. 1996). Our approach is currently based on modeling elevation gradients, which are correlated with temperature (Keleher and Rahel 1996). Our attempts to delineate the amount and distribution of suitable habitat (i.e., patch structure) for salmonids have relied on empirical relationships between downstream distribution limits of juveniles and elevation or geographic gradients (see also Flebbe 1994). Our work has been with Lahontan cutthroat trout in the eastern Lahontan Basin in southeast Oregon and northern Nevada, and bull trout in the upper Boise River Basin in southern Idaho (Fig. 26.2).

Delineation of patches for bull trout in the Boise River Basin relied on information from surveys of juvenile distributions, which suggested a sharp increase in occurrence above an elevation of 1,600 meters (Rieman and McIntyre 1995; Dunham and Rieman 1999). This distribution limit was used to delineate the amount and distribution of suitable habitat patches within the basin. In the case of Lahontan cutthroat trout, a geographic model was necessary to account for changes in the elevation of distribution limits over the eastern Lahontan Basin, which covers a much broader area (Dunham et al. 1999). Geographic

location (latitude and longitude) explained over 70 percent of the variation in the elevation of downstream distribution limits for Lahontan cutthroat trout.

Patch delineation involved linking models of downstream distribution limits with a geographic information system (GIS). Predicted downstream distribution limits were used to delineate the size and distribution of watersheds with suitable habitat. We defined patches of suitable habitat as the watershed area upstream of predicted elevations for downstream distribution limits. Defining patches in terms of watershed area is consistent with the view that watershed characteristics have an important influence on stream habitats (Montgomery and Buffington 1998). Local or regional variation in watershed characteristics may have an important influence on the development of stream channels and aquatic habitat (Burt 1992), and patch structuring and patterns of species occurrence may vary accordingly.

An alternative, and perhaps more precise, measure of patch size would be actual length of stream occupied within a watershed. Length of stream occupied requires information on both up- and downstream distribution limits, whereas watershed area requires only information on downstream distribution limits. Stream length might be important where there is strong local variability in climate and geomorphology, or when barriers to fish movement within streams limit upstream distributions. If barriers are important, fish may only be able to occupy a very limited amount of habitat, and patch sizes estimated by stream length and watershed area could differ substantially. Limited evidence suggests the influence of barriers on fish distributions within streams is generally minimal, though important exceptions do exist (e.g., Kruse et al. 1997; Dunham et al. 1999).

Another potentially important localized factor is occurrence of nonnative trout. In the case of Lahontan cutthroat trout, for example, downstream distribution limits were significantly restricted when nonnative trout were present (Dunham et al. 1999). This effect was not consistent or predictable, so we could not simply account for the effect of nonnative trout in defining distribution limits and patch sizes. Earlier models of occurrence of Lahontan cutthroat trout did not detect an effect of nonnative trout (Dunham et al.

1997), but this study did not provide clear definition of patch structure.

Because localized factors within streams (e.g., geomorphic features, nonnative fish) may place constraints on the amount of habitat that can be occupied, patch areas may not reflect the "effective" size of habitat available to fish. To remedy this potential problem, we examine model interaction terms and prediction errors for streams with and without known constraints. The alternative is to directly map local features of stream habitats across large areas to delineate patches, which is often difficult to justify with limited resources.

Modeling Species Occurrence

Delineation of patches and patch structuring within drainage networks provides a template for predicting species occurrence. Our approach is essentially a two-tiered model: (1) define distribution limits in terms of geographic or elevation gradients to delineate suitable habitat patches, (2) predict occurrence of fish within patches. To predict species occurrence, we have focused on influences of the geometry of patches within landscapes, namely the size, isolation, and spatial distribution of patches (see Rieman and Dunham 2000). Here, we focus on patch size. Patch size may be related to fish occurrence because habitats in larger patches may be more complex and resilient to disturbance and should generally support larger populations.

Rieman and McIntyre (1995) used multiple logistic regression to model occurrence of bull trout in the Boise River Basin and found patch area to be the strongest predictor. Other significant factors included patch isolation and road density within patches (Dunham and Rieman 1999). Solar radiation and occurrence of nonnative brook trout were not associated with occurrence, and patterns of occurrence were not spatially aggregated (Dunham and Rieman 1999). Logistic regression of occurrence of Lahontan cutthroat trout in relation to patch area revealed a highly significant ($P < 0.0001$) and positive relationship (J. B. Dunham unpublished data).

Although analyses of data for Lahontan cutthroat trout are preliminary, some interesting common themes are suggested by the results. For both species,

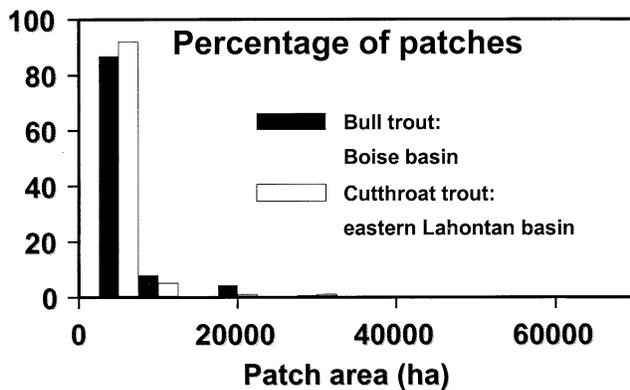


Figure 26.3. Comparison of patch size distributions for bull trout (*Salvelinus confluentus*) in the upper Boise River Basin and Lahontan cutthroat trout (*Oncorhynchus clarki henshawi*) in the eastern Lahontan Basin.

patch area appears to be a significant correlate of species occurrence. This is a common pattern emerging for many species in both terrestrial and aquatic ecosystems (Bender et al. 1998; Moilanen and Hanski 1998; Magnuson et al. 1998; Hanski 1999) and a general prediction from island biogeography and metapopulation theory (Hanski and Simberloff 1997). More interesting are the details of the relationship between patch size and species occurrence.

First, an examination of patch size distributions reveals that size distributions are remarkably similar for both species and skewed toward very small patches (Fig. 26.3). This means that very few patches are likely to have a high probability of occurrence and that a few large patches may be very important for both species. In terms of total area of potential habitat occupied, bull trout in the Boise River Basin occupy relatively more (46 percent) than Lahontan cutthroat trout in the eastern Lahontan Basin (36 percent). When the actual responses of both species to changes in patch size are compared (Fig. 26.4), it is clear that both species are likely to occur when patch sizes exceed about 10^4 ha in area. Relative to Lahontan cutthroat trout, bull trout are more likely to occur in smaller patches (Fig. 26.4), which may explain why bull trout occupy a larger percentage of suitable habitat overall.

Because a common definition for patches was used for both species, we were able to compare specific responses of each to variability in patch size. In the future, analysis of occurrence in relation to other char-

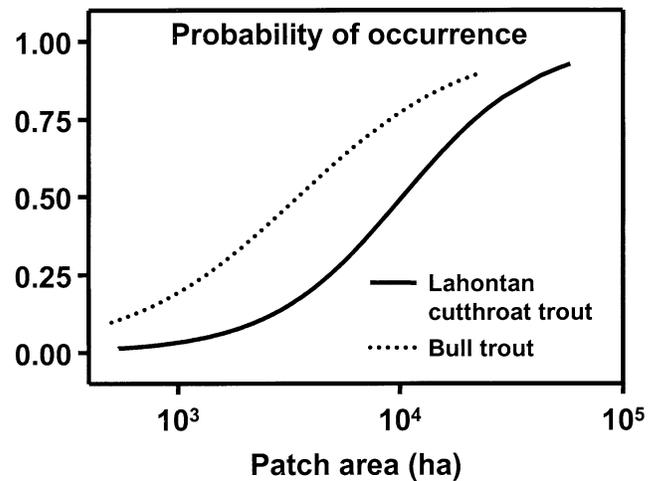


Figure 26.4. Predicted probability of occurrence in relation to patch size (area) for bull trout (*Salvelinus confluentus*) in the upper Boise River Basin and Lahontan cutthroat trout (*Oncorhynchus clarki henshawi*) in the eastern Lahontan Basin.

acteristics of patches may reveal additional insights. Although the biology of these two species differs in important ways (Rieman and Dunham 2000), these results provide general themes to guide efforts to conserve and manage these species, along with important details relevant to particular species or environments they inhabit.

Evaluating Model Prediction

The most relevant measure of a classifier is its expected error rate (EER, Lachenbruch 1975). Among possible EER estimators, leave-one-out cross-validation is a nearly unbiased estimator of out-of-sample model performance (Fukunaga and Kessel 1971) that provides a measure of overall predictive ability without excessive variance (Efron 1983). Leave-one-out cross-validation involved removal of an individual observation from the data set, fitting a model with the remaining observations, and predicting the omitted observation. Model probabilities of occurrence greater than or equal to 0.50 were classified as predicted occurrences. Fits between observed and predicted occurrences were summarized as classification error rates, summarized over all observations and by response.

Because the results for Lahontan cutthroat trout were preliminary, we focused on classification (omission) and prediction (commission) errors for the

model of bull trout occurrence reported by Dunham and Rieman (1999). Based on the overall classification error rate (Table 26.1), our logistic regression model was fairly accurate with a 19.7 percent error rate. This suggests a good fit between the data and our logistic regression model, but this simple measure of predictive ability does not reveal insights into potential bias or sources of error. When modeling species occurrences, biological responses are usually approximated assuming some predefined statistical distribution. For example, logistic regression assumes a binomial response distribution and a logit link. Hence, model accuracy is likely a function of how faithfully the distribution approximates the biological response. To illustrate, we fit the bull trout occurrence data to a k-nearest-neighbor (KNN) model, a relatively flexible nonparametric classification technique that does not require distribution assumptions or a strong assumption implicit in specifying a link function (Hand 1982). The overall KNN error rate was 17.3 percent, only slightly lower than the logistic regression model, which suggested a reasonable fit of the data to the assumed binomial distribution. Category-wise (i.e., response-specific) classification and prediction error rates can also be influenced by response-specific sample size (Agresti 1990). In general, error rates are higher for less-frequent responses. For instance, bull trout occurrence had the lowest sample size and the highest prediction and classification error rates for both the logistic regression and KNN models (Table 26.1). Additionally, logit model error rates are influenced by the choice of the baseline category (e.g., modeling presence or absence, Agresti 1990).

Choice of statistical model may be important, but

error in determination of occurrence can also bias model predictions. For example, bull trout in smaller patches may also occur at lower densities, which may affect probability of detecting fish. Another important point is that predictions from the model are not precise. Lower and upper confidence intervals for slope estimates of the patch area effect range from 48 to 61 percent of the point estimate (see Dunham and Rieman 1999). We do not have enough confidence in the precision of our models to believe that model “accuracy” can be reasonably assessed by analysis of errors of omission or commission alone. In other words, there is a need for both statistical and biological “validation” of models. We suspect similar limitations apply to many other models of species occurrence.

Implications for Models of Species Occurrence

The efficacy of a patch-based approach depends on how clearly patches can be defined. Even if the definition of patch structuring is clear, it may not be realistic to treat patches as independent of the landscape in which they are embedded. The degree to which the landscape “matrix” should be considered in models of species occurrence will depend on patch contrast, aggregation, and scale of study (Kotliar and Wiens 1990; Wiens 1996a,b), as well as life history characteristics of the species in question. For example, many salmonids have complex migratory behaviors, and interactions between migratory behavior and habitat outside of spawning and rearing areas (e.g., Fig. 26.1) may affect species occurrence within patches (Rieman and Dunham 2000).

A patch-based approach to modeling patterns of species occurrence has several important advantages. In the simplest sense, a patch-based approach permits stratification of models of species occurrence. In purely statistical terms, stratification may be a useful tactic to increase the precision of model predictions. Lack of consideration or knowledge of patch structure can produce mismatched inferences between patterns of occurrence and habitat characteristics. Many common approaches to subdividing landscapes into pixels, polygons, political boundaries, and so forth

TABLE 26.1.

Summary of leave-one-out cross-validation classification and prediction^a error rates^b for bull trout patch occupancy models.

Patch status	<i>n</i>	Logistic regression	K ^c -nearest neighbor
Occupied	29	0.276 (0.276)	0.207 (0.258)
Unoccupied	52	0.154 (0.154)	0.154 (0.120)

^aPrediction in parenthesis.

^bOmission and commission errors, respectively.

^c*k* = 12 nearest neighbors.

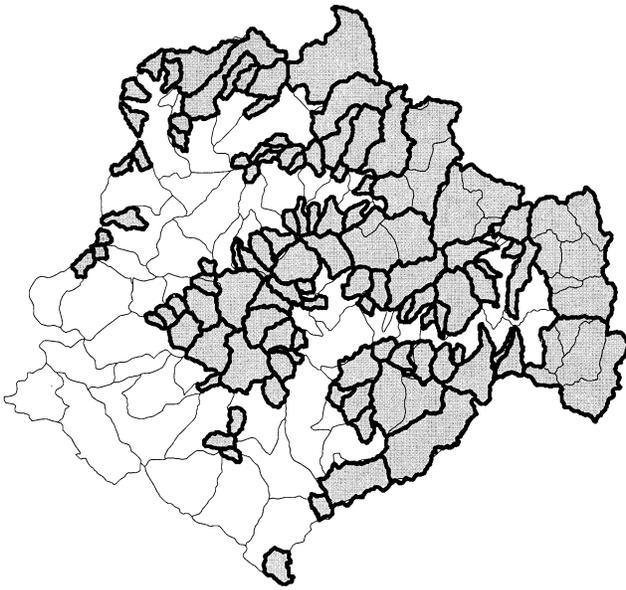


Figure 26.5. Watersheds are often subdivided into hydrologic units (HUs) for classification and analysis of aquatic habitats (Maxwell et al. 1995). This overlay of sixth-field HUs (thin lines) and patches (heavy lines with shading) for bull trout (*Salvelinus confluentus*) in the Boise River Basin shows that patch and HU watershed boundaries can be substantially different.

may not adequately reflect patch structuring (e.g., Fig. 26.5).

Patch Structure and Scale

Patch structure can have important implications for finer-scale models of occurrence. Often, fish-habitat relationships for salmonids are considered within relatively small sites ranging from individual pools and riffles to stream segments (10⁰–10²-meter) (Fausch et al. 1988; Angermeier et al., Chapter 46). At these spatial scales, occurrence of fish among sites is probably not independent, because they are nested within a larger area supporting a population that is influenced by larger-scale environmental variation and fish movement (Schlosser 1995; Gowan and Fausch 1996b).

At finer scales (e.g., stream segments within patches), the existence of suitable but unoccupied habitat is a possibility, especially when patch dynamics are characterized by extinction and/or (re)colonization (Rieman and Dunham 2000). This implies that absence of fish at sites may not be a function of site-

scale habitat quality, but rather a characteristic of a larger patch (e.g., size, isolation), within which sites are nested. As hypothesized by Kotliar and Wiens (1990), larger-scale environmental variation may place important constraints on patterns nested at smaller scales. Nonspatial models of species occurrence implicitly assume organisms are free to select all habitats, but this is not true if external constraints (e.g., spatial isolation, dispersal barriers) or internal constraints (e.g., homing, philopatry) are important (see Rosenberg and McKelvey 1999 for a recent example).

At larger ecological scales, patches supporting local populations may be aggregated within landscapes, perhaps forming metapopulations or “semi-independent networks” (Hanski 1999; Rieman and Dunham 2000). For salmonids in streams, patterns of hydrologic connectivity may produce spatially aggregated clusters or “networks” of patches. Whether defined as metapopulations or otherwise, aggregates of patches supporting local populations can be characterized by a number of larger-scale characteristics, such as number of patches, size distribution, isolation, land form, land type, or climatic associations. Models linking species occurrence to aggregate characteristics may be considered in terms of occurrence of a single or multiple local populations.

Management Applications

One of the strongest motivations for better predictive models of species occurrence is the need for better data to support conservation planning for threatened and endangered species (the two species described here are listed as threatened under the U.S. Endangered Species Act). Key steps in conservation planning include (1) delineation of units for conservation, (2) risk or status assessment of the units, and (3) prioritization of management and species recovery actions. Delineation of biotic units and selection of appropriate biological responses and/or criteria have been the focus of much debate, especially for salmonid fishes (e.g., Nielsen 1995; see also Paetkau 1999). Patch-based models provide a useful context for integrating different elements of biological diversity (e.g., compositional, structural, functional; Franklin 1988) that should be considered in delineation of conservation

units at different scales (Noss 1990b). In terms of risk or status assessments, patch-based models provide managers with a useful description of the amount and distribution of suitable habitat (e.g., patch size, isolation). Predictions from models of occurrence can provide important information to address fundamental questions about habitat conservation (e.g., How many habitats? How large? Where to focus?). Other patch attributes (e.g., occurrence of other species, land cover, ownership, climatic data, etc.) can also be added to prioritize species recovery actions.

Patch-based models have provided useful insights for a wide variety of species (see Hanski 1999). Ignoring potential patch structure or applying overly simplistic or unrealistic habitat classifications may result in models with little biological relevance or poor pre-

dictive power. The past several years have seen a dramatic increase in our capacity to generate predictor variables (e.g., via GIS, online databases, etc.) and better analytical models to predict species occurrence. As our ability in these important areas increases, however, we should not lose sight of the biological responses we wish to understand *and* predict.

Acknowledgments

Thanks to Debby Myers for GIS support. The Biological Resources Research Center, University of Nevada, Reno, provided additional support. Constructive reviews by Paul Angermeier, Erica Fleishman, Kirk Krueger, and three anonymous reviewers improved earlier versions of the manuscript.