

Nonparametric Multiplicative Regression for Habitat Modeling

Bruce McCune

Oregon State University, Corvallis, Oregon U. S. A.

8 July, 2011

© B. McCune, 2005-2009, 2011

Suggested citation: McCune, B. 2011. Nonparametric Multiplicative Regression for Habitat Modeling.
<<http://www.pcord.com/NPMRintro.pdf>>.

Table of Contents

| | |
|--|-----------|
| Introduction..... | 3 |
| What are Habitat Models?..... | 3 |
| What is a Niche?..... | 3 |
| Kinds of Habitat Models | 5 |
| General approaches | 5 |
| Model types | 6 |
| A new approach..... | 7 |
| Global and local models | 9 |
| Species responses to single habitat factors..... | 10 |
| Appropriate response types | 10 |
| Binary or quantitative | 10 |
| Binary or quantitative | 10 |
| Quantitative | 10 |
| Binary | 10 |
| Combining response shapes | 12 |
| Summary of general habitat modeling strategy | 14 |
| Nonparametric Multiplicative Regression (NPMR)..... | 15 |
| Design principles for habitat modeling with NPMR..... | 15 |
| Basic concepts | 15 |
| Notation | 16 |
| Forms of NPMR | 16 |
| SpOcc - NPMR..... | 16 |
| LM - NPMR | 17 |
| LLR – NPMR..... | 20 |
| LLogR - NPMR..... | 21 |
| Model Building..... | 21 |
| Calibration vs. application..... | 21 |
| Model specification | 21 |
| Model evaluation..... | 21 |
| Model selection | 26 |
| Model application..... | 28 |

| | |
|---|-----------|
| What to report for NPMR..... | 30 |
| Example Fits for the Gaussian-Gaussian Problem..... | 31 |
| The data set..... | 31 |
| Least squares models..... | 31 |
| NPMR models..... | 33 |
| Detailed Example of NPMR for a Tiny Data Set..... | 35 |
| Comparison of NPMR in HyperNiche to Nonparametric Multiple Regression in S-Plus..... | 41 |
| Example of Overfitting..... | 43 |
| Appendix A – Notation..... | 45 |
| Appendix B – Method for Free Search..... | 46 |
| Appendix C – Example Publications Using NPMR and HyperNiche..... | 49 |
| References..... | 54 |

Introduction

What are Habitat Models?

Defined broadly, a habitat model represents a relationship between a species and factors that control its existence. The performance of a species in relation to those factors can be measured in numerous ways, most commonly presence-absence, abundance, physiological rates, and demographic rates. Habitat factors in the narrow sense are variables that describe the characteristics of particular locations – the habitat. In a broader sense, habitat factors include all other determinants of species performance, including other species, disease, disturbance history (including time since disturbance), conditions at the moment of a physiological measurement, or even time of day. This broader sense is easy to work with in a practical way. Statistical tools that are good at empirically modeling of habitat in the narrow sense also work well for many other predictors of species performance.

Conceptual habitat models have been formative in ecological theory (Austin et al. 1994). Consider, for example, the Hutchinsonian niche, an n -dimensional hypervolume (Hutchinson 1957, 1965; see below), and Whittaker's influential diagrams of species responses to environmental gradients (e.g. Whittaker 1956). Despite the central importance of species response functions, our understanding of these is surprisingly primitive, particularly when considering more than one factor at once.

Our difficulties in understanding species' response functions derive in part from the inability of standard statistical models to represent them well. This document describes that problem and proposes a workable, flexible solution: nonparametric multiplicative regression (NPMR). This is a method that can be used to represent species response surfaces in a multidimensional niche space.

What is a Niche?

The word “niche” has been both a convenience and a frustration to ecologists. The convenience is that on some level, all ecologists can communicate with this word. The frustration is the difficulty and multiplicity of precise definition (Wiens 1989).

Do not despair: the Hutchinsonian niche (1957; Fig. 1) provides a simple, practical basis for putting the concept of niche to work. I suspect that few readers will have actually read Hutchinson's precise and elegant original definition of niche, so here it is verbatim:

Consider two independent environmental variables x_1 and x_2 which can be measured along ordinary rectangular coordinates. Let the limiting values permitting a species S_1 to survive and reproduce be respectively x'_1, x''_1 for x_1 and x'_2, x''_2 for x_2 . An area is thus defined, each point of which corresponds to a possible environmental state permitting the species to exist indefinitely. If the variables are independent in their action on the species we may regard this area as the rectangle ($x_1 = x'_1, x_1 = x''_1, x_2 = x'_2, x_2 = x''_2$), but failing such independence the area will exist whatever the shape of its sides.

We may now introduce another variable x_3 and obtain a volume, and then further variables $x_4 \dots x_n$ until all of the ecological factors relative to S_1 have been considered. In this way an n -dimensional hypervolume is defined, every point in which corresponds to a state of the environment which would permit the species S_1 to exist indefinitely. For any species S_1 , this hypervolume N_1 will be called the fundamental niche of S_1 . Similarly for a second species S_2 the fundamental niche will be a similarly defined hypervolume N_2 .

It will be apparent that if this procedure could be carried out, all X_n variables, both physical and biological, being considered, the fundamental niche of any species will completely define its ecological properties. The fundamental niche defined in this way is merely an abstract formalisation of what is usually meant by an ecological niche.

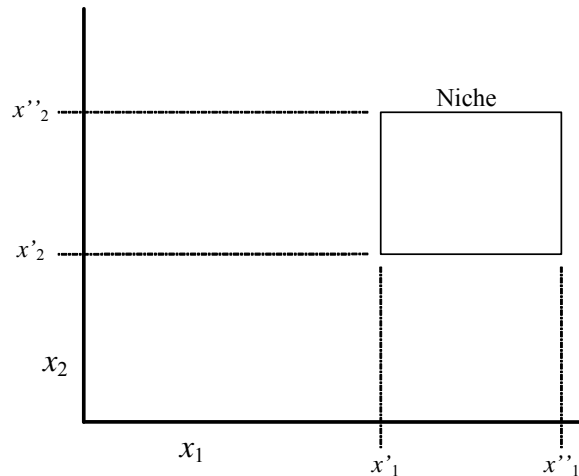


Figure 1. The Hutchinsonian niche. The axes x_1 and x_2 are selected dimensions of the niche space.

Hutchinson (1957) later in the same paper pointed out four limitations to this definition of niche. Each of these limitations is paraphrased below, along with some parenthetical comments on the importance of these limitations to the statistical modeling of niche proposed here:

1. The definition assumes equal probability of persistence in all parts of the niche, when in fact conditions will become suboptimal toward the boundaries of the niche. [This limitation is easily remedied by modeling probability of occurrence or abundance as continuous response variables throughout the n -dimensional space.]
2. All environmental variables can be linearly ordered. [In practice, both quantitative and categorical variables can be included as predictors of species occurrence or abundance.]
3. The model refers to a single instant of time. [Hutchinson gives the example of two animals that eat the same foods, but one at night and one in the day. Curiously, he does not seem to consider using time-of-day (or other temporal variables) as dimensions of the niche space. From the standpoint of statistical modeling, there is no reason why time-of-day or day-night could not be used simply as additional dimensions of the niche space.]
4. “Only a few species are to be considered at once, so that abstraction of these makes little difference to the whole community.” [This he resolves in the same paragraph, pointing out that additional species can be regarded as part of the coordinate system.]

Hutchinson’s concept of the niche has been criticized because of how he connected his concept of niche to the community as a level of organization (Morrison & Hall 2002). If this is a problem, it is because of what has been built on top of the concept of niche, rather than anything in definition itself. The practicality of the concept has also been questioned, because it invokes an unlimited number of dimensions (Peters 1991). But the concept has an easy practical manifestation -- operationalizing the niche demands selection of a finite number of dimensions for a model. There is no need to pretend or assume that all dimensions of the niche are represented in a particular model. However, the fact that not all important factors are represented in a model has important statistical consequences (Kaiser et al. 1994).

From a modeling standpoint, we should be able to accept a broad concept of the niche, where the dimensions can include physical environmental factors, other species or other biological factors, resource gradients, or time (e.g. time since disturbance) – any of the factors that impinge on the survival of an organism. Selection of those dimensions defines the utility and meaning of a particular model.

Kinds of Habitat Models

General approaches

In representing species response functions, ecologists usually use simplistic statistical models that cannot capture the nonlinear multifactor relationships of a species to its habitat. The models usually lack interaction terms and the default response shapes are typically linear (as in multiple linear regression; Fig. 2) or sigmoid (as in logistic regression). Yet most or all ecologists accept the concept that species commonly have hump-shaped response functions to environmental gradients (Fig. 2). Furthermore, we expect the shape of this response to depend on other factors. In other words, factor interactions should be expected in our models.

Linear models (Fig. 2) may be appropriate in some cases, such as species responses to short environmental gradients. Likewise, logistic response functions are sometimes appropriate, for example, with a sigmoid relationship between a species probability of occurrence and a successional gradient. Many other possibilities exist, however.

The standard ecological concept for the relationship of species to environmental gradients is a unimodal, hump-shaped curve, such as those popularized by Whittaker (Fig. 2). Though widely accepted as a theoretical model, where are the statistical models of single-species response functions that incorporate a unimodal response? These are rare in the ecological literature, (but see Huisman et al. (1993) for Gaussian logistic regression). Even more rare are models where the shape of the unimodal response depends on another variable, yet this should be the norm.

Many different nonlinear response functions are possible. For example, consider a weedy biennial plant species that colonizes bare soil in the first year after a fire. A rosette is produced the first year, followed by flowering and death the second year. If occupancy of the site prevents colonization in years after the first, then the species disappears after the second year. The response of the abundance of this biennial to the temporal gradient thus forms a step function (Fig. 3).

McCune & Grace (2002) summarized the history of our efforts to describe species responses to single environmental gradients:

Describing species responses to environmental gradients is fundamental to developing and testing ecological theory, improving methods of community analysis, improving our use of indicator species in environmental assessments, predicting geographical and environmental distributions of species from sample surveys, and predicting the impacts of climate change on vegetation (Austin et al. 1994). How can we best represent these responses mathematically? Investigators have tried smoothing functions (Austin 1987), generalized linear modeling with third-order polynomials (Austin et al. 1990), beta functions (Austin et al. 1994, but see Oksanen 1997), maximum likelihood with a Gaussian response model (Oksanen et al. 1988), least squares with a Gaussian response model (op. cit.), weighted averaging (op. cit.), and logistic regression (Huisman et al. 1993). The complexity of the problem has defied a general satisfactory solution. Response curves are often skewed, sometimes polymodal, and the species optimum often lies outside the sampled range. In many cases, a Gaussian model is not appropriate.

Our simplistic statistical models – multiple linear regression, logistic regression, or some other form of GLMs – do not readily accommodate complex multiplicative combinations of hump-shaped and other nonlinear responses.

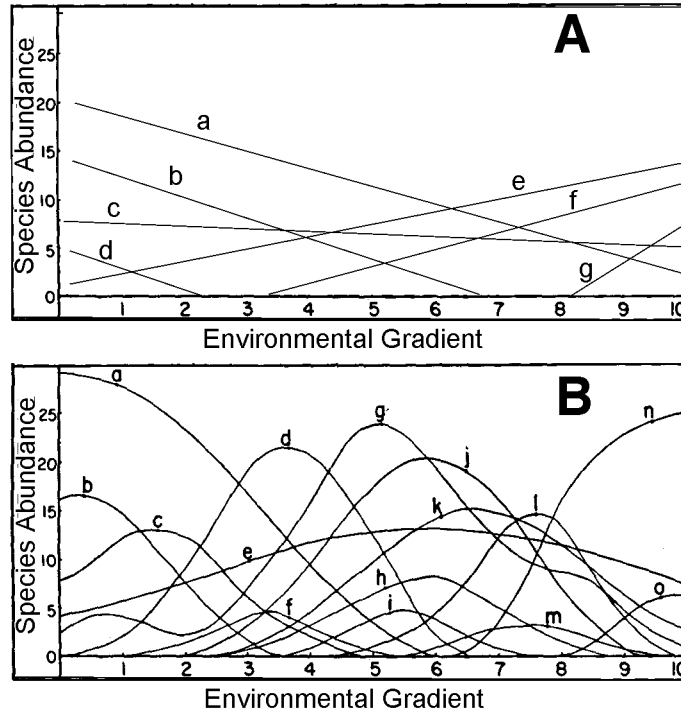


Figure 2. Hypothetical responses of species abundance to an environmental gradient. Lettered curves represent different species. **A. Linear** responses. **B. Hump-shaped** responses (after Whittaker 1954).

Austin (2002) described three components of statistical modeling in ecology: an ecological model, a data model, and a statistical model. “The ecological model consists of the ecological knowledge and theory to be used or tested in the study. The data model consists of the decisions regarding how the data are collected and how the data will be measured or estimated. The statistical model involves the choice of statistical method, error function and significance tests.”

Using Austin’s terminology, ecologists’ statistical models often mismatch their ecological models. As discussed above, the usual statistical models are too rigid and inflexible to accommodate well the complex response surfaces that by all accounts must be present. We need to understand the consequences of that mismatch and we need to develop ways of avoiding that mismatch in the future.

With parametric modeling, “Assumptions about the shape of the response of species to an environmental variable (usually termed an environmental gradient) are central to any predictive modeling effort” (Austin 2002). Nonparametric multiplicative regression negates this statement: predictive modeling can be effective without making any assumptions about the shapes of species responses to ecological factors.

Model types

The following short list of the kinds of habitat models is not comprehensive, but attempts to summarize some of the main modeling approaches. See Guisan and Zimmerman (2000) for a more comprehensive summary and Scott et al. (2002) for an extended treatment of the problems of habitat modeling with numerous examples.

Linear models and generalized linear models (GLMs) in one form or another are the most commonly used methods for habitat models. Most ecologists have used linear and logistic regression for habitat modeling (e.g. Boyce et al. 2002; Franklin 1995; Guisan & Zimmermann 2000; Scott et al. 2002; Fleishman et al. 2003). Reliance on linear models has hampered the development of ecology. Huston (2002) provides examples of how the use of linear models has obfuscated even simple ecological relationships, such as changes in diversity along gradients in productivity.

Generalized additive models (GAMs; Hastie 1990) are becoming increasingly popular for habitat modeling (e.g., Bio et al. 1998; Bowman and Azzalini 1997; Erjnaes 2000; Heegaard 2002a, 2002b; Leathwick 1995; Yee & Mitchell 1991; Zaniwski et al. 2002). GAMs fit individual terms with smooth nonparametric functions (rather than, for example, linear or polynomial functions). This makes an important step by avoiding the assumption of a particular response shape (or order of the polynomial).

By "term" we mean an additive effect in a model. So, for example, say you had a model with latitude and longitude as predictors. You could fit the GAM for latitude and longitude separately: $y = m_1(\text{Lat}) + m_2(\text{Long}) + \text{constant}$. This model has two terms plus the constant. Each of the terms is an unspecified smooth function m of one of the predictors. Alternatively, you could fit a single smooth function m_{12} of both predictors simultaneously. The model $y = m_{12}(\text{Lat}, \text{Long}) + \text{constant}$ has one term (plus the constant).

Thus a GAM can combine predictors multiplicatively (e.g., combining latitude and longitude as predictors in a single smooth function), but individual terms are combined additively. The multiplicative models advocated here could be considered a special case of GAMs with a single term containing multiplicative weights across all predictors.

Classification trees have been used as habitat models, using presence-absence as a binary response variable (e.g. Austin et al. 1994, Franklin 1998; Vayssières et al. 2000). In this context, classification recursively partitions an environmental space so as to maximize the homogeneity of the response variable within subspaces. Classification trees have the advantage of automatically modeling interactions among predictors. They operate, however, within the constraint of using only perpendicular planes to divide the environmental space into subspaces. Regression trees are similar in concept and limitations but apply to continuous response variables.

Ecological Niche Factor Analysis (ENFA; Hirzel et al. 2002, Zaniwski et al. 2002) models habitat based on the special case of having presence data only. This method generates "pseudo" absences by various techniques.

Rule-based methods (including GARP, Genetic Algorithm for Rule-set Prediction; Peterson 2001, Peterson et al. 2002, Peterson & Robins 2003; Scachetti-Pereira 2002) seek combinations of habitat-based rules to optimize prediction of species presence-absence. These methods can also incorporate traditional statistical models, such as logistic regression.

Multivariate adaptive regression splines (MARS; Friedman 1991, Friedman & Roosen 1995). MARS selects predictors and partitions the predictor space by fitting spline functions in multiple dimensions. Use of tensor product spline basis functions accommodates interacting predictors, making this a potentially powerful method for ecological habitat models. As a recursive partitioning method, MARS is related to CART, differing in that MARS produces a continuous model.

A new approach

Our understanding of species responses to habitat factors has been stifled by an emphasis on forcing responses into molds defined by parametric models. More rapid progress surely can be made by adopting a more open exploratory approach. Scott (1992, p. 3) emphasized the importance of this exploratory approach:

There is a natural flow among the parametric, exploratory, and nonparametric procedures that represents a rational approach to statistical data analysis. Begin with a fully exploratory point of view in order to obtain an overview of the data. If a probabilistic structure is present, estimate that structure nonparametrically and explore it visually. Finally if a linear model appears adequate, adopt a fully parametric approach. Each step

conceptually represents a willingness to more strongly smooth the raw data, finally reducing the dimension of the solution to a handful of interesting parameters.

Ecologists have most often worked in the opposite order, beginning by adopting a linear model (e.g. ANOVA or multiple regression). Alternative structures are seldom sought.

The problem is particularly difficult for habitat models, because we know that species rarely respond linearly to individual factors. This much is clear from studies of species responses to single ecological factors (see above). The problem is compounded with multiple habitat factors. What do species responses look like for multiple habitat factors considered simultaneously?

The challenge for habitat modeling is exactly the same as that expressed for data analysis in general by Scott (1992, p. 5): “The modern challenge in data analysis is to be able to cope with whatever complexities may be intrinsic to the data. The data may, for example, be strongly non-Normal, fall onto a nonlinear subspace, exhibit multiple modes, or be asymmetric [all of these are commonly true of species responses]. Dealing with these features becomes exponentially more difficult as the dimensionality of the data increases, a phenomenon known as the *curse of dimensionality*.”

This curse applies to models of species performance in relation to multiple factors. As the number of factors increases, the number of potential interactions increases exponentially. The number of transformations or combinations of transformations similarly inflates. The number of possible combinations of factors to include or exclude balloons.

“An incorrectly specified parametric model has a bias that cannot be removed by large samples alone... The modern emphasis on robust estimation correctly sacrifices a small percentage of parametric optimality in order to achieve greater insensitivity to model misspecification” (Scott 1992, p. 33).

In habitat modeling, ecologists tend to have vague, often unstated, ideas about model form. Traditionally ecologists have visualized only one, or at most two, habitat factors at a time. Ecological theory has not come close to dictating an appropriate form for species response to multiple habitat factors.

Despite the profusion of books and papers in the statistical literature on **smoothing** techniques and **nonparametric regression** (Bowman & Azzalini 1997, Eubank 1999, Fan & Gijbels 1996, Green & Silverman 1994, Hastie et al. 2001, Scott 1992, Wand and Jones 1995), these techniques have largely been ignored by ecologists. Yet some use of smoothing functions for more than one simultaneous predictor exists in the literature (Gignac et al. 1991a; Gignac et al. 1991b; Huntley et al. 1989, 1995). Gignac et al. (1991a, b) generated 3-D response surfaces for species abundance along environmental gradients. The response surfaces were generated from gridded abundance data, using distance-weighted means. The use of distance-weighted smoothing functions is allied to NPMR. Their approach had several drawbacks: an arbitrarily selected (rather than optimized) search radius, arbitrary ways of dealing with zeros and outliers, and no method for cross-validation.

Locally-weighted smoothing (or regression) using the LOWESS approach was applied to habitat models in two and three dimensions by Huntley et al. (1989, 1995). Their approach shares some features with our use of NPMR, but differs in several important respects: (1) limitation to two or three pre-selected predictors instead of a conducting a free search for the best model using an indefinite number of variables from a pool of available predictors, (2) they did not optimize the breadth of the smoothing function for each predictor, instead choosing it arbitrarily, and (3) they fit their model at fixed intervals within the plane of predictors, with linear interpolation between intervals, rather than fitting the model for each data point.

Nonparametric multiplicative regression (NPMR) as used here is based on kernel functions to weight observations. This is a smoothing technique that can be cross-validated and applied in a predictive way. Many other smoothing techniques are well known, for example smoothing splines and wavelets. Optimum choice of a smoothing method depends on the specific application. A key advantage to using NPMR for habitat modeling is that the approach is easily extended to many dimensions (predictors). The multidimensionality is provided multiplicatively – this automatically and parsimoniously models the complex interactions among predictors. This flexibility comes at a price: computational speed – optimizing the selection of predictors in a multiplicative model is a computationally intensive process. Although other model types will continue to be useful in habitat modeling, NPMR can improve both the quality of model predictions and the simplicity of model construction. NPMR can be applied to either presence-absence or quantitative response data.

Global and local models

To place nonparametric regression into a common framework with normal least-squares regression, we need to differentiate between a local model and a global model. A global model is a relationship that applies throughout the sample space. A local model is fit to a particular region of the space, but the model can differ in different regions of the sample space.

With simple linear regression, the global model is a straight line relationship throughout the whole sample space (Fig. 3). The model is fit to the whole data set simultaneously, so there is no local model. Every point is given equal weight in the analysis, so the local weighting function is flat (Fig. 3).

Similarly logistic regression has no local model and a flat weighting function (Fig. 3), but it differs from simple linear regression in that the global model is a sigmoid curve of probabilities in relationship to predictors. Gaussian logistic regression is like the usual logistic regression, except that a different link function specifies a hump-shaped relationship to the predictors instead of a sigmoid relationship (Fig. 3).

Nonparametric regression with a kernel function takes the opposite approach. Instead of rigidly specifying a global model, the global model is not specified and can take any form. But we now introduce a local model, a relationship that is fit to each data point, weighting data points according to their distance from the target point (Fig. 4). The weighting function specifies the way in which those weights vary with distance from the target point, and is sometimes known as the “kernel.”

In the simplest form, the “SpOcc” model specifies a rectangular weighting function – also known as a uniform or boxcar kernel. Data points within a window receive full weight (1.0) and data points outside the window receive no weight (0.0; Fig. 4). The local model is flat, in that we simply average the observed values within the window, without assuming any trend within the window.

If instead of a rectangular window, we give less and less weight to observations that are increasingly distance from the target point, then we can use a Gaussian (or bell-shaped) weighting function (Fig. 4). The local model can be flat (i.e. simply calculating a local weighted mean; LM-NPMR) or we can fit for each point a trend. The trend can be locally linear (LLR-NPMR) or locally logistic (LLogR-NPMR; Fig. 4).

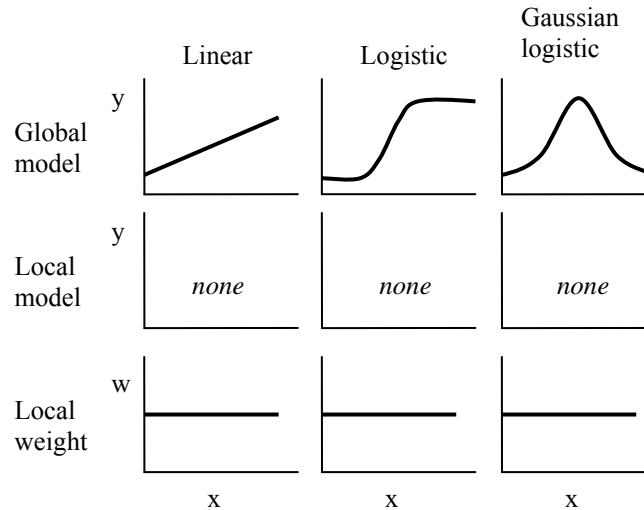


Figure 3. Some common parametric habitat models. These models specify an overall relationship without reference to a local model. Contrast this with nonparametric regression (below) with a local model but no global model.

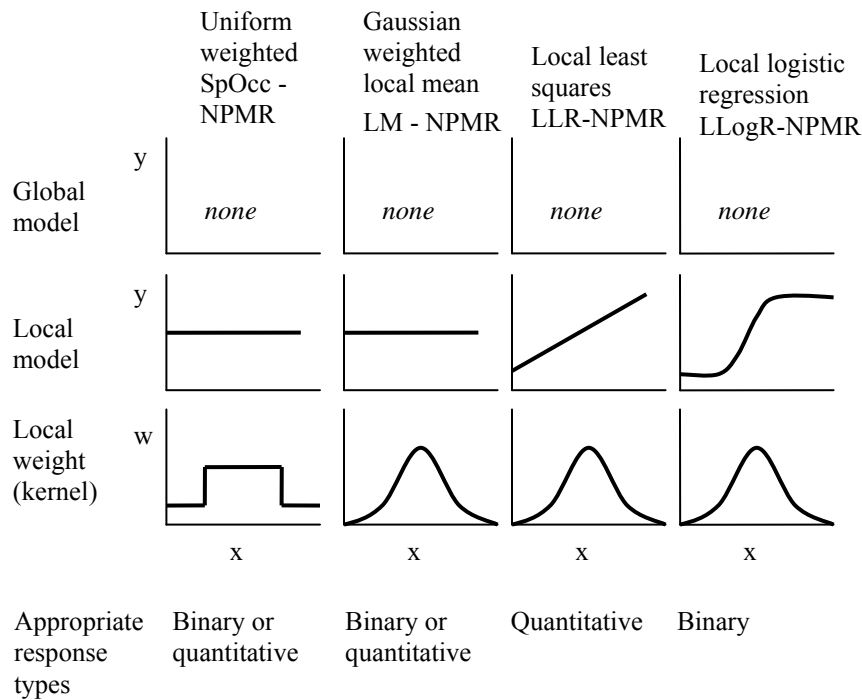


Figure 4. The global model, local model, and weighting function used in several forms of nonparametric regression.

Species responses to single habitat factors

As a starting point for habitat modeling, consider a short catalog of fundamentally different kinds of species response to habitat factors (Fig. 5). These response shapes are not at all controversial – most ecologists could think of or accept an example of each curve shape.

The **hump-shaped** or unimodal response (Fig. 5) was popularized by R. H. Whittaker and reproduced in numerous textbooks. A hump-shaped response is implicit in Shelford's law of tolerance and the Hutchinsonian niche. One example application in my area of interest is the prevalence of the epiphytic lichen *Lobaria oregana* at middle elevations in the Cascade Range in Oregon, U.S.A. The species is abundant at about 400-900 m and absent below 100 m and above about 1300 m. Presumably this reflects the underlying physiological tolerances to the climatic gradients that occur with elevation.

A **linear** response (Fig. 5), or nearly so, is readily conceivable when we measure species performance over a short gradient. Continuing the previous example, a linear approximation of *Lobaria oregana* with elevation is adequate over a short elevational span, say 100-200 m.

Abundance of *Lobaria oregana* has a **sigmoid** relationship (Fig. 5) to time since stand initiation in even-aged forests. Even in its optimal habitat, the population initially builds very slowly, but after about 100-250 years the biomass builds rapidly, finally reaching a plateau in old-growth forests. A sigmoid response may also be reasonable for studies of probability of occurrence based on presence-absence data, particularly when the habitat factor is measured over a short range. Over long environmental gradients the likelihood function is more likely to be unimodal.

A species that resides near a surface may decline in a **negative exponential** way with distance from the surface (Fig. 5). For example, phytoplankton inhabiting the ocean surface might decline rapidly in abundance with increasing depth, approximating a negative exponential function. Or, a pioneer species that cannot persist through time might decline in a negative exponential way. An immediate pulse of establishment after disturbance may be followed by gradual loss from the community. For example, weedy

shade-intolerant plants such as *Taraxacum* and *Hypochaeris* frequently colonize disturbed soil in clearcuts. These species rapidly decline as the canopy closes, but persist longer in canopy gaps.

Sudden loss or gain of a species can be represented by a **step function** (Fig. 5). This might be caused by an extreme environmental event, catastrophic disease, or disturbance. Step changes can result from any source of a sudden synchronized establishment or mortality. For example, weedy biennial plants establishing immediately after a disturbance will be vegetative the first year, produce seed the second year, then die. If, in the meantime, the habitat has become fully occupied, establishment of the biennial might fail until after further disturbance.

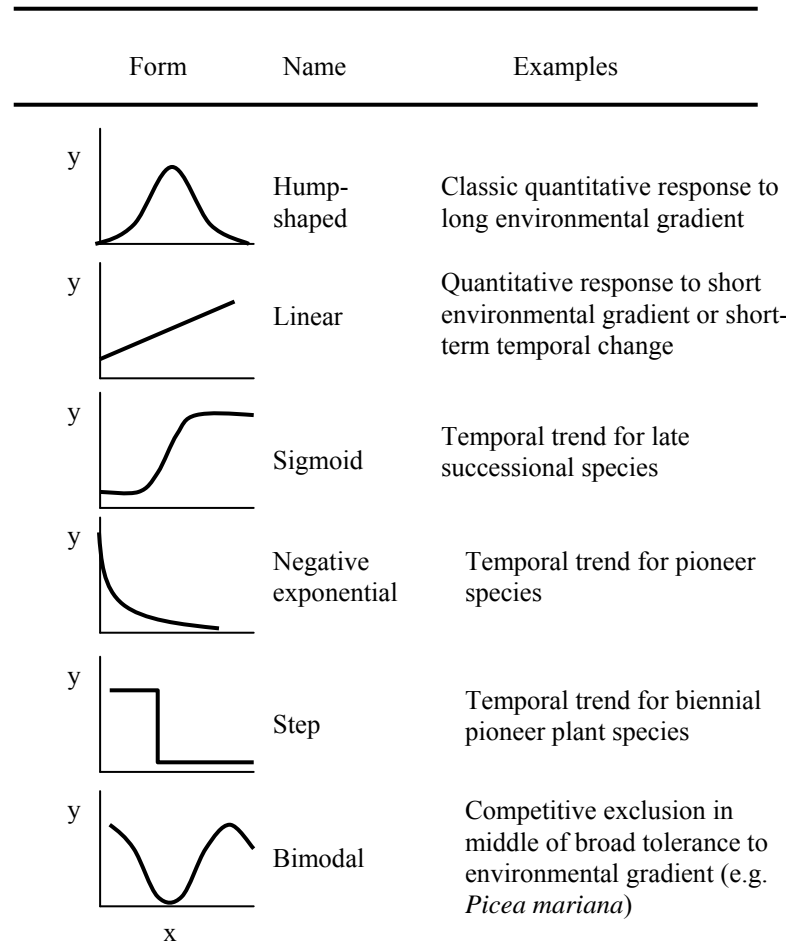


Figure 5. Example shapes for species responses to environmental and temporal gradients. These can be viewed as responses in terms of abundance, probabilities of occurrence, demographic variables, or physiological rates.

McCune & Grace (2002) gave the following example of a **bimodal** species response (Fig. 5). “The distributions of black spruce (*Picea mariana*) and eastern redcedar (*Thuja occidentalis*) along a moisture gradient in parts of the American boreal forest are classic examples of bimodal species distributions (Curtis 1959, Loucks 1962). Black spruce occurs in soil pockets on dry rock outcrops as well as in wet *Sphagnum* bogs. In intermediate (mesic) sites with deeper, well-drained soils, black spruce is often a minor species or absent. These sites are dominated by white spruce (*Picea glauca*), sugar maple

(*Acer saccharum*), or balsam fir (*Abies balsamea*).” Although the fundamental niche of black spruce may have a unimodal response to moisture, its realized niche is a more complex shape because its performance is disrupted by competition with other tree species (McCune 2006). Species interactions are likely to produce complex responses to abiotic environmental factors.

Although we can readily envision all of these response shapes, our statistical models tend to fall back on a very limited subset of these, primarily linear models. The next section describes some of the statistical tools used to represent species response functions.

Combining response shapes

What happens when we combine simple species responses to multiple habitat factors? Consider the following hypothetical example, say that y is the reproductive rate of a population, x_1 is food, and x_2 is shelter. The model

$$y = b_1x_1 + b_2x_2 + b_0$$

says that the reproductive rate is increased by the availability of food and shelter, and that increasing either of these alone can increase the reproductive rate. But consider a population in the Antarctic. The simple additive model comes to the erroneous consequence that you will have a fairly high reproductive rate if you give them lots of food but no shelter. Likewise, the model errs in concluding that the population will reproduce if you give them shelter but no food. This can be partially corrected with an interaction term, but this prescribes a particular shape across the whole response surface, a portion of a hyperbolic paraboloid (Fig. 6).

The additive model is shown in Fig. 6. When x_1 and x_2 are combined additively, the response surface forms a plane tilted in the 3D space formed by the response variable and the two predictors. By adding a multiplicative (interaction) term, the model becomes

$$y = b_1x_1 + b_2x_2 + b_3x_1x_2 + b_0$$

The response surface can now be curved (Fig. 6), even though each term in the model is linear. As dimensions (predictors) are added, however, the problem rapidly becomes difficult to manage, and non-parsimonious, because one must consider adding two-way and higher order interactions.

The alternative lies in multiplicative models, where the effect of each variable can depend on the value of other variables. One approach is to adapt nonparametric curve fitting techniques, the components being combined multiplicatively rather than additively – this is nonparametric multiplicative regression.

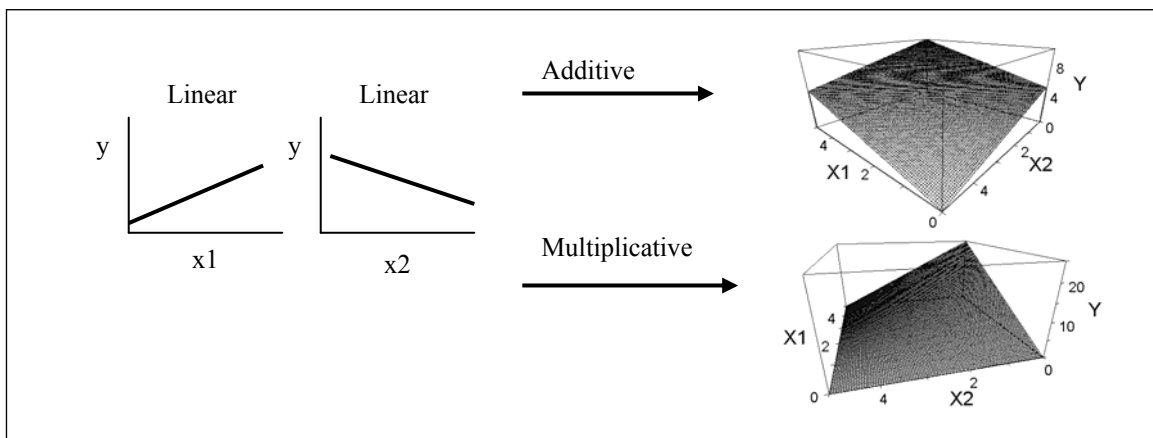


Figure 6. Combining **linear** species responses, y , to two habitat factors, x_1 and x_2 . The surface from the multiplicative model (i.e. with an interaction term) is a portion of a hyperbolic paraboloid.

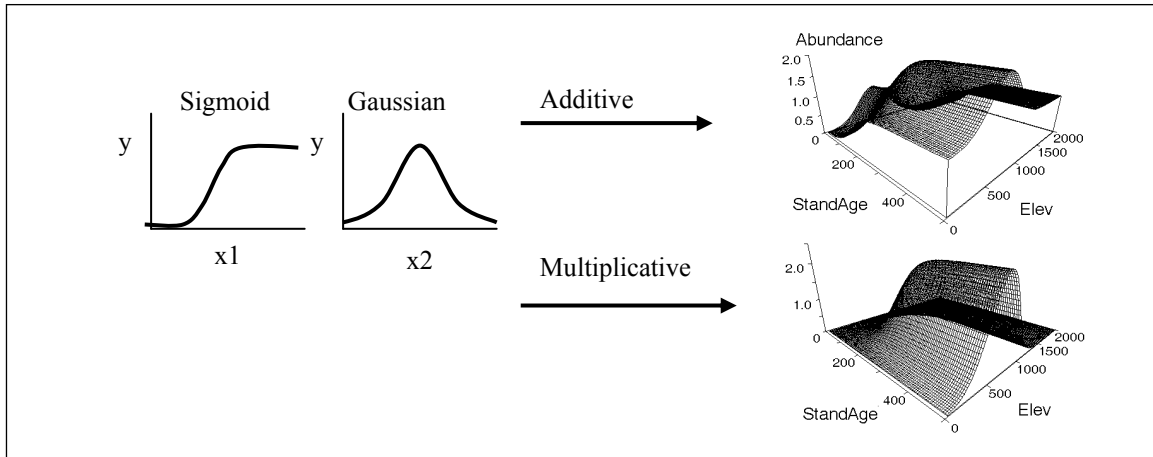


Figure 7. Combining **nonlinear** species responses y to two habitat factors, x_1 and x_2 .

The two-factor problem can be made more realistic by combining two nonlinear shapes, for example sigmoid and Gaussian curves (Fig. 7). This example corresponds to the previous example of the abundance of the epiphytic lichen, *Lobaria oregana* (y), in response to time since disturbance and elevation. In the absence of disturbance, *Lobaria* has a sigmoid increase through time (x_1). The species has a unimodal response to elevation (x_2), peaking between 400 and 900 m.

An additive model of the two nonlinear functions could be:

$$y = \frac{e^{-a_1+b_1x_1}}{1 + e^{-a_1+b_1x_1}} + e^{-a_2*(x_2-b_2)^2}$$

The first term is the sigmoid response to time since disturbance and the second term is the Gaussian response to elevation. A strictly additive model means that if one factor is highly favorable and another factor highly unfavorable, the model predicts a moderate abundance. But the basic law of life and death is multiplicative, not additive: if any one factor is lethal then no level of any other independent factor can compensate for it. The response shape from the additive model (Fig. 7) is, therefore, fundamentally flawed. For example, the additive model yields high abundance when x_1 is high (old forests) at both very high and very low values of x_2 (elevation). In nature, if elevation is unfavorable (very high or very low elevations), then no *Lobaria* is found, no matter how old the forest.

Assuming that we know the shape of the response function, we could, in theory, develop a multiplicative model that would better fit the data (lower right part of Fig. 7). In practice, however, we need multifactor models for which we cannot even visualize the response surface, simply because ecological theory does not inform us as to response shapes in the multidimensional spaces implicit in multifactor models. Furthermore, we need a parsimonious and effective method for building multifactor models. For these reasons, we turn to a model fitting procedure that automatically and effectively represents the interactions among predictors without needing to see the response surface in the hyperspace of habitat factors.

Despite the obvious importance of interacting predictors in ecology, multiplicative nonlinear habitat models are actually very rare in ecology. A few examples are Martinez-Taberner et al. (1992) and Huntley et al. (1989, 1995). While these methods provided insights, they were not built into a general, flexible modeling framework that could be applied to a wide range of problems.

Summary of general habitat modeling strategy

Choice of a habitat modeling technique depends on many factors. Using the flowchart (Fig. 8) may help you decide what general class of models is appropriate for a particular problem.

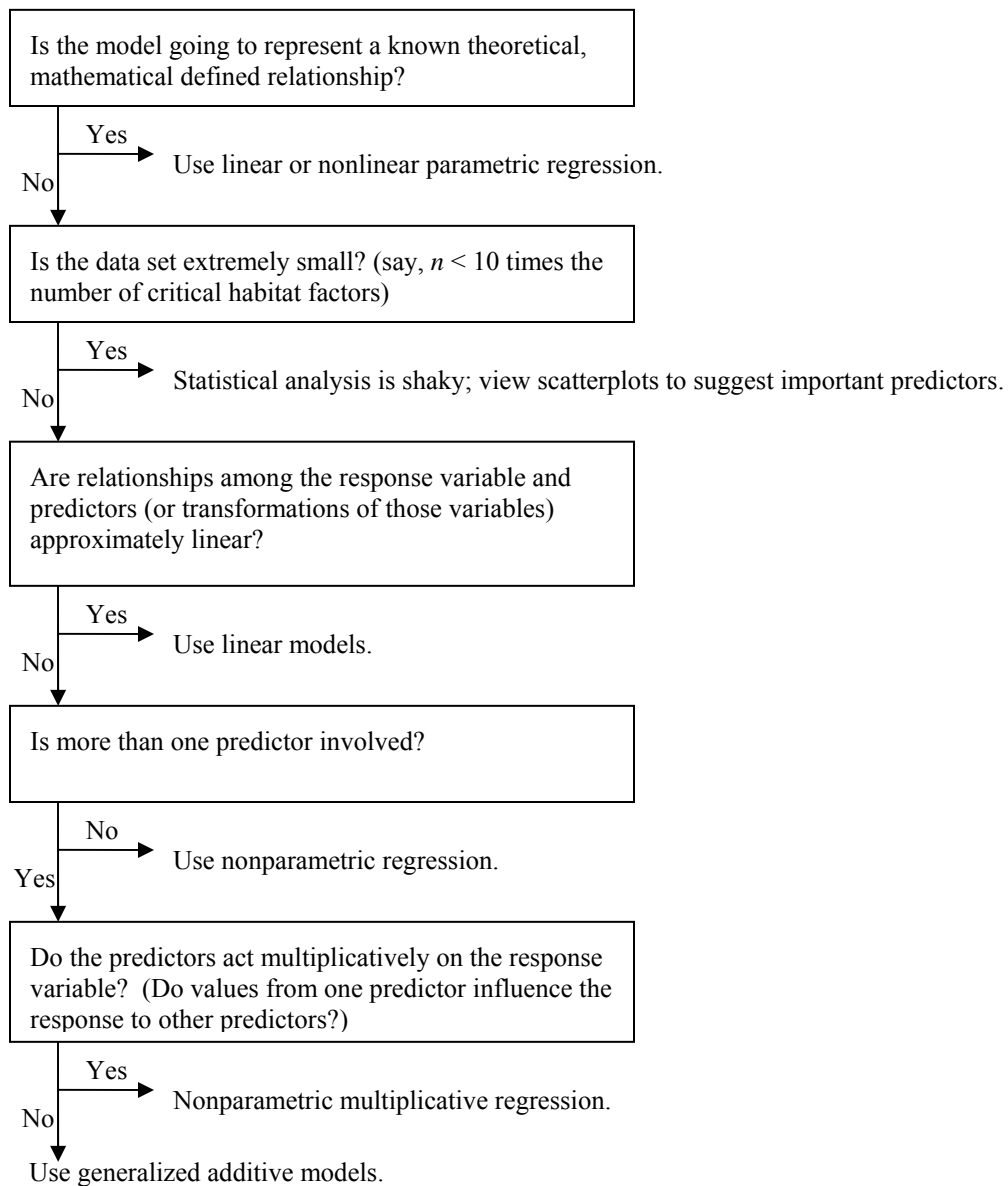


Figure 8. Decision tree for some general classes of habitat models.

Nonparametric Multiplicative Regression (NPMR)

Design principles for habitat modeling with NPMR

Nonparametric Multiplicative Regression (NPMR) is a class of statistical techniques with a number of variants. The software package HyperNiche (McCune & Mefford 2004) applies the concepts of NPMR to predictive habitat modeling and species response functions in particular. Designing a special form of NPMR for habitat modeling has allowed us to incorporate many features that help us with this particular modeling problem. The most important characteristics of the form of NPMR used in HyperNiche are:

- Use a method for evaluating model quality that can be applied to any habitat model. For **presence-absence (binary) data**, the log likelihood ratio (we use $\log_{10}B$) evaluates predictive ability compared to a standard naïve model. The naïve model estimates the probability of occurrence at a particular point as the overall frequency of the species in the data. We apply this with a leave-one-out cross-validation to help guard against overfitting the model. For **quantitative data**, the naïve model estimates the abundance of a species at a particular point as the mean abundance of the species in the whole data set. We evaluate the model quality with a cross-validated R^2 value (xR^2).
- Allow both categorical and quantitative predictors.
- Always use multiplicative weighting so that variables are combined multiplicatively rather than additively.
- Variable selection is based on a cross-validation procedure to reduce problems of overfitting.
- Variable selection allows both addition and deletion of variables and simultaneous adjustment of the tolerances of the weighting function. Explore these possibilities with a stepwise or comprehensive search.
- Optimize the kernel function by using a cross-validation procedure to select smoothing parameters.
- Allow input/output of gridded GIS data, specifically input of predictor grids and output of grids of species estimates (e.g., abundance or probability of occurrence).

Basic concepts

Nonparametric regression, like linear regression, seeks relationships between a response variable and one or more predictors. Nonparametric regression does not, however, seek coefficients in a mathematical equation of fixed overall form. Instead, it seeks to optimize fit to the data without reference to a specific global model.

What kinds of surfaces can be represented by NPMR? Any function – a smooth 1:1 mapping between a response variable and a set of predictors – can, in theory, be detected by NPMR. NPMR cannot describe surfaces that have overhangs, discontinuities, or cusps.

To use NPMR one must choose a local model and a kernel function (Table 1). The local model specifies the shape of the function that is used to fit a value for a specific point in the space defined by the independent variables. The kernel specifies a weighting, in essence specifying how local is “local”. In HyperNiche, the kernel function is optimized by using a cross-validation procedure to select a weighting parameter.

We discuss each of four basic model forms (Table 1). At present only the first three are included in HyperNiche.

Table 1. Four forms of nonparametric multiplicative regression used for habitat modeling.

| Model name | Local model | Weighting around each target point (kernel) | Response data types |
|--------------|---------------------------|---|------------------------|
| SpOcc - NPMR | local mean | uniform within a window | binary or quantitative |
| LM – NPMR | local mean | Gaussian | binary or quantitative |
| LLR – NPMR | local linear regression | Gaussian | quantitative |
| LLogR - NPMR | local logistic regression | Gaussian | binary |

Notation

A full listing of notation is given in Appendix A. Some of this is repeated here:

\mathbf{X} = matrix of predictors (habitat or environmental variables) with $i = 1$ to n rows (sample units) and $j = 1$ to m columns (variables).

\mathbf{y} = vector of observed presence-absence, abundance, or other response variable. This is a column vector of $i = 1$ to n rows (sample units).

s_j = standard deviation of the Gaussian weighting function for predictor variable j , applied to a given predictor such that the full range of observed values for that variable falls over 6 standard deviations.

\mathbf{v} = vector specifying the habitat at the target point, this vector being row vector of $j = 1$ to m columns (variables).

w^*_{ij} = weight applied to point i for predictor j . The asterisk indicates that it is a univariate weight, as opposed to a weight from the matrix \mathbf{W} .

\hat{y}_v = fitted value or estimated probability of occurrence of species at target point v .

Forms of NPMR

SpOcc - NPMR

The concepts of the species occurrence model were initially based on Peterson (2000), then developed further and applied by McCune et al. (2003). Peterson dubbed the procedure “SpOcc,” a name with trekkie appeal which stands for “Species Occurrence Modeler.” The model was implemented for binary data as an unpublished add-in module “SpOcc” to PC-ORD 4 (McCune & Mefford 1999), then later generalized to the NPMR add-in module. Since then it has been incorporated into a more general stand-alone program for habitat modeling, HyperNiche (McCune & Mefford 2004).

With binary data, SpOcc uses the proportion of a species’ occurrence in an environmental neighborhood to estimate the likelihood of the species occurring at a target site (Fig. 9). The environmental neighborhood consists of sites nearby in an multi-dimensional environmental space, the space defined by values for one or more environmental or habitat variables. It can thus be considered an “environmental neighborhood model” for predicting species occurrences. The method requires information on known (sampled) sites, including the presence of species at the sites (presence/absence or abundance), and the environmental characteristics (or other predictors) at the sites. To estimate the species occurrence for a new site (the target site), the model applies data from sites that lie close to the target site in the n -dimensional environmental space (the environmental neighborhood).

An environmental neighborhood can be defined in two different ways. Most simply, the neighborhood can be defined by a “window” with sharp edges (Figs. 9, 10). This is the method used by the “SpOcc” form of NPMR (Peterson 2000, McCune et al. 2003). The size of the window is defined by a tolerance range around the target site. All observations within the window are given equal weight (one), while all observations outside the window are given zero weight. This is, in essence, nonparametric regression using one of the simplest kernel approaches, the *local mean estimator* (Bowman and Azzalini 1997, p. 49), but using an unconventional uniform weighting function.

Alternatively, the neighborhood can be defined without sharp boundaries, rather weighting observations near the target point more strongly than observations that are distant from the target point. Again, the concept of “distance” applies to an environmental space, not necessarily geographic space. Weights range from near zero to one. This is the method described below under “LM-NPMR”.

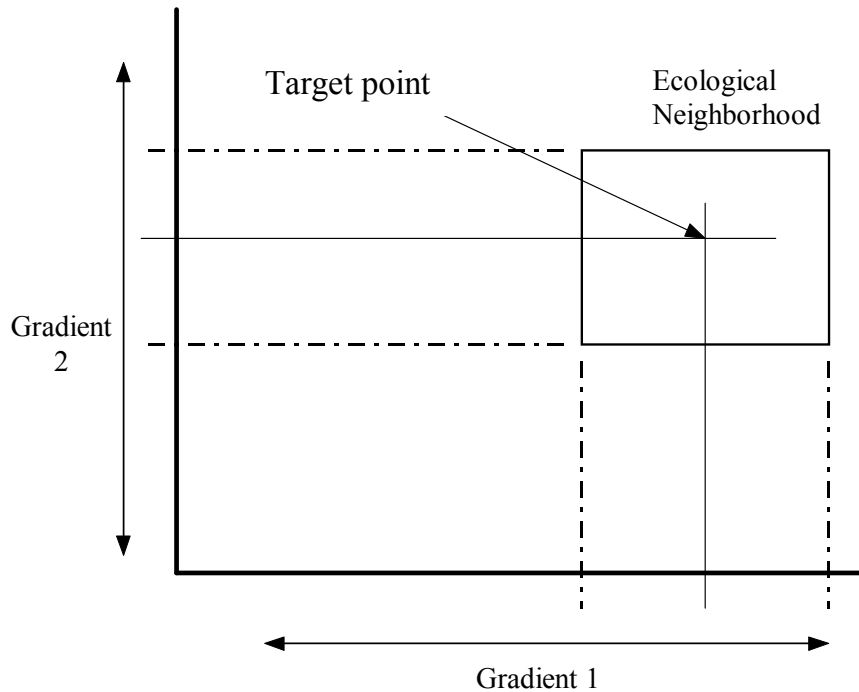


Figure 9. Two-dimensional representation of the Species Occurrence Model (SpOcc-NPMR). An estimation of the abundance or likelihood of species presence is sought for the target point. This estimate is based on the frequency of the species in the ecological neighborhood, defined by segments of the two gradients.

Response curves fitted with the SpOcc form of NPMR (i.e. with a uniform weighting function) tend to be choppy than those with a Gaussian weighting function. The sharp edge of the ecological neighborhood tends to produce fitted surfaces with sharp changes, owing to the influential behavior of individual points falling in or out of the window. We recommend, therefore, using another form of NPMR with a Gaussian weighting function, except where one anticipates more precipitous changes in the response variable.

LM - NPMR

The relationship between weight given to an observation and distance from the target point can be defined in many different ways. A simple, flexible solution is to use a Gaussian (hump-shaped) function centered on the target point (Fig. 10). A observation with exactly the same environment as the target point would receive full weight (1.0), smoothly diminishing to near zero weight with increasing distance from the target point. In contrast, with the window method (Fig. 10), weights are, in essence only zero or one in a square function (Fig. 10). How rapidly the weights diminish with distance from the target point can be

tuned with the smoothing parameter, in this case the standard deviation of the Gaussian curve (Fig. 11). Selecting a large standard deviation is comparable to having a broad window; conversely a small standard deviation gives appreciable weight only to observations that are very similar to the target point.

We apply the term “tolerance” to both the half-width of the square window and the standard deviation of a Gaussian weighting function. In both cases the word is biologically apt, because a species with narrow tolerance to a habitat factor will have a smaller window, whether specified as a square, hard-edged window or a window with edges made fuzzy by a Gaussian function.

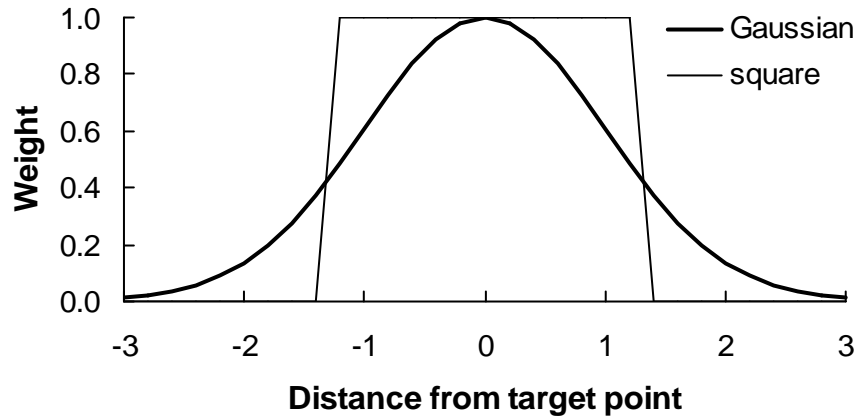


Figure 10. Uniform (square) weighting function versus Gaussian function. In this example, the full range of the environmental variable spans 6 unspecified units and the target point is in the center of that span.

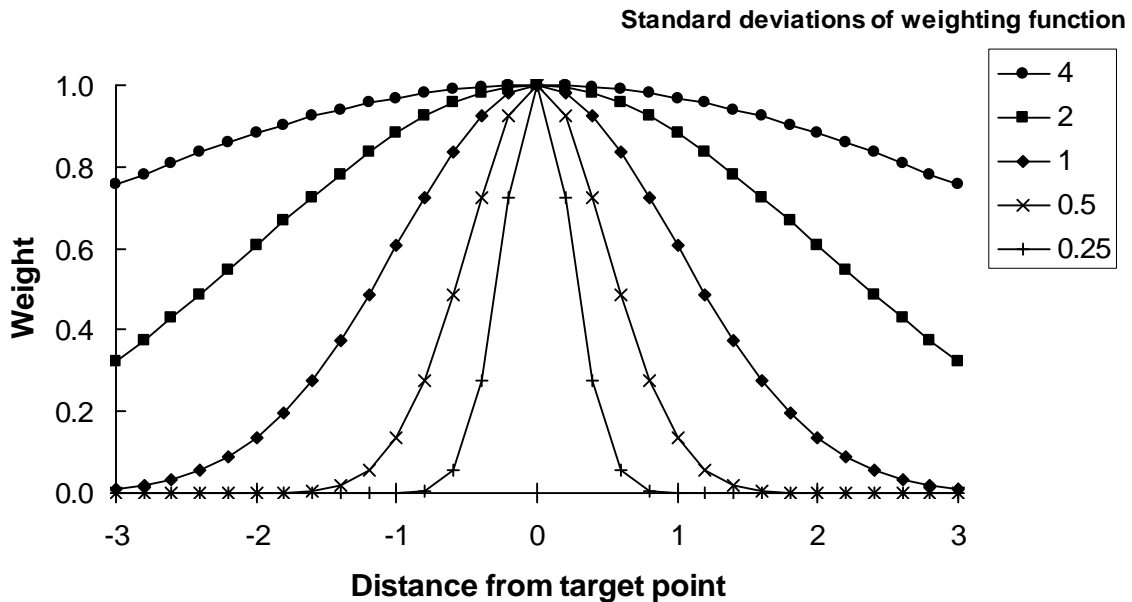


Figure 11. Gaussian function for weighting observations relative to distance from an observation to the target point in environmental space. The standard deviation of the Gaussian function controls how quickly weights diminish with distance from the target point. In this example, the full range of the environmental variable spans 6 unspecified units and the target point is in the center of that span.

The Gaussian function shown (Fig. 11) is altered from its usual probability density function, such that the height of the peak always equals one and the area under the curve is no longer equal to one. If s is the standard deviation, x is the value of an environmental variable, and w is the resulting weight, then:

$$w = e^{-\frac{1}{2}[(x - \bar{x}) / s]^2}$$

With more than one environmental factor in the model, the weights for individual factors can be combined multiplicatively into a single weight for a given observation. See for example, the equation for two dimensional density estimation and nonparametric regression in Bowman and Azzalini (1997, pages 6 and 53, respectively). Then, for a given point those weights are used in a weighted average on the observations of presence/absence in the data set.

The weight applied at a particular point, relative to the target point \mathbf{v} , is:

$$w^*_{ij} = e^{-\frac{1}{2}[(x_{ij} - v_j) / s_j]^2}$$

The asterisk with the w indicates that this is a univariate weight (the weight for a single predictor j at sample unit i).

We then combine weights across environmental variables, multiply the combined weight by the observed value y , and divide by the sum of the combined weights to give a weighted average as an estimate of the probability of occurrence at target point \mathbf{v} :

$$\hat{y}_v = \frac{\sum_{i=1, i \neq v}^n y_i \left(\prod_{j=1}^m w^*_{ij} \right)}{\sum_{i=1, i \neq v}^n \left(\prod_{j=1}^m w^*_{ij} \right)}$$

This is the local mean estimator of Bowman and Azzalini (1997, p. 49) extended multiplicatively to m dimensions. The notation $i \neq v$ indicates that if the target point v is one of the calibration data points, then it is excluded from the basis for the estimate of y_v . This is the crux of the built-in cross-validation procedure. Cross-validation has a long history of use in nonparametric regression (e.g. Clark 1975; Hardle et al. 1988, 1992; Hardle 1990; Assaid & Birch 2000).

The NPMR equation above should become the standard equation for exploratory habitat models. It is a fundamental equation in ecology, because it explicitly recognizes that **ecological factors combine multiplicatively**. This equation confronts ecological complexity head-on by recognizing that **we must build the big structure from estimates of the local structure**, rather than assuming a simple form for the big structure. Species response surfaces are complex, nonlinear shapes that are not well recognized by traditional parametric models.

For every estimate of the response variable one can calculate a neighborhood size (n_i^*), the amount of data bearing on that particular estimate:

$$n_i^* = \sum_{i=1}^n \left(\prod_{j=1}^m w^*_{ij} \right)$$

where $0 < n_i^* \leq n$. If $n_i^* = 0$, then no estimate is possible for that point. Setting a minimum n_i^* required for an estimate protects against estimating a response in a region of the predictor space with insufficient data.

With binary (presence-absence) data, to consider the estimates of y as probabilities for the occurrence of species, the sampling of sites must be random or randomized within strata used as predictors in the model. Even when this requirement is not met with strict randomization, the output is still useful for relative comparisons among sample units.

Estimates for sample points near the ends of gradients are weakened by the absence of data for the environmental neighborhood that lies just beyond the end of the gradient. This decreases the accuracy of the model near the ends of the gradient, because estimates are made on less data. It also introduces a slight bias near the ends of the gradient, making predictions more similar to the sampled portion of the gradient than expected.

LLR – NPMR

A fault of the local mean estimator is that estimates near the ends of the predictors are biased toward the central tendency of the response variable (Bowman & Azzalini p. 50-51). This occurs because the closer a target point is to the edge of the sample space, the less data are available beyond the target point. This bias can be removed by using a local linear estimator rather than a local mean estimator. The local linear estimator is simply a weighted least squares problem, the weights provided by the kernel function such that points near the target point receive more weight than points far from the target point.

LLR was proposed as early as Cleveland (1979) and developed further by Fan and Gijbels (1992) and Fan (1993). Some characteristics of LLR are:

- Bias is reduced near the edges of the data set.
- As the kernel function becomes broad, the fitted curve will smoothly approaches traditional least squares regression while the local mean smoothly approaches a horizontal line parallel to the predictor axis with an intercept equal to the global mean.

The local linear estimator can be represented in matrix notation if we first create the design matrix \mathbf{Z} containing the transformed predictors plus a first column of 1s. The predictors are transformed by subtracting each value for a given variable from the corresponding value for the target point. \mathbf{Z} has n rows $m+1$ columns (variables). The i th row of \mathbf{Z} has the elements:

$$[1 \quad (x_{1i} - v_1) \quad (x_{2i} - v_2) \quad \dots \quad (x_{ji} - v_j)]$$

We also create a $n \times n$ diagonal matrix of weights, \mathbf{W} . The i th diagonal element of \mathbf{W} is the product of the weights for all variables $j = 1$ to m . For sample unit i , the diagonal element is

$$w_{ii} = \prod_{j=1}^m w_{ij}^*$$

Then the local linear estimator is the first element of the weighted least squares solution, \mathbf{b} :

$$\mathbf{b} = (\mathbf{Z}'\mathbf{W}\mathbf{Z})^{-1} \mathbf{Z}'\mathbf{W}\mathbf{y}$$

This regression equation is solved for each target point, so a data set with $N=500$ will require 500 weighted least-squares regressions, for every set of trial tolerances.

Despite the improved fit to responses near the ends of the ranges of predictors, local linear models are less conservative, and can produce wild estimates under some circumstances. They are less conservative, because estimates of the dependent variable can be outside of its observed range. This behavior can be particularly noticeable and offensive with small data sets. In contrast, the local mean can never produce an estimate outside the observed range of the dependent variable. In practice, the decision between LM- and LLR-NPMR is a tradeoff between avoiding the known bias near the edge of the sample space and avoiding the possibility of wild estimates in that same region. If the former is the more serious risk, then use LLR-NPMR, while if the latter is possible, use LM-NPMR. In general, small data sets are more safely modeled with LM-NPMR.

LLogR - NPMR

A third approach, not yet implemented in HyperNiche, uses local logistic regression to estimate the likelihood of occurrence. The method applies to binary response variables. The local model is a standard logistic regression, except that observations are weighted by their distances to the target point (Bowman and Azzalini 1997, Equation 3.4, p. 55) and weights should be combined multiplicatively across variables. The advantage of using an estimator based on local logistic regression over a local mean estimator is that the latter is biased at the edges of the predictor space. Because the target point lies near the edge of a habitat variable, the estimate is biased toward the overall frequency of occurrence. Using local regression (linear or logistic) removes this bias (op cit., p. 50).

Model Building

Calibration vs. application

Habitat modeling generally has two phases, calibration and application. In the **calibration phase** we use empirical data on species abundance or species presence-absence to evaluate the model's ability to estimate abundance or likelihood of occurrence from the independent variables. The calibration phase is used to (1) decide which variables to include, (2) select a window size or smoothing parameter (tolerance) for quantitative variables, and (3) indicate how much we can trust predictions from the model in the application phase. In the **application phase** (see Model Application below), we estimate abundance or likelihood of occurrence for sites at which species occurrence or abundance is unknown, based on the values of the predictors. When we estimate the likelihood of occurrence from presence-absence data rather than predict presence/absence per se, we call this "estimating" or "forecasting" rather than "prediction."

Model specification

If an NPMR model does not yield an equation with coefficients, what is the form of the model? In its simplest form an NPMR model is completely specified by the following items:

1. the data set used in the calibration phase,
2. statement of the local model (e.g., local mean, local linear, local logistic),
3. statement of the weighting function (e.g., Gaussian or uniform),
4. a list of one or more independent variables,
5. specification of whether the independent variables are treated as categorical or quantitative, and
6. a tolerance or smoothing parameter for each quantitative variable. If windows have sharp edges, then this range specifies the exact width of the window. If windows have fuzzy edges, then the specified tolerance is used as one standard deviation in the Gaussian-weighted window.

Model evaluation

We use the term "evaluation" for the process of analyzing the predictive success of models, following Oreskes et al. (1994) and Guisan and Zimmerman (2000). Others have called this "validation" or "accuracy assessment." HyperNiche automatically uses a leave-one-out form of cross validation. It is applied during the search for the best model, so that choice of predictors and their tolerances are based on the results of cross-validation.

Binary response data

We sought a method for evaluating model performance that could be applied to any method of estimating likelihood of occurrence. We considered only methods that avoid the arbitrary conversion of continuous estimates of probability of occurrence into a statement of "present" or "absent" (Fleishman et al. 2003). One common method of evaluating the performance is a pseudo-*R*-squared statistic (Agresti 1990, pp. 110-112):

$$\text{pseudo } R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2}$$

where \hat{y}_i is the fitted value for the model and y_i is the observed value (1 or 0). This can be applied to the results of any model that predicts values of 1 or 0. Although this is the same form as a traditional R^2 value, we call it “pseudo” because applied in this context it has no fixed lower bound. Negative values are possible when the “sum of squares regression” (the numerator) is larger than the total sum of squares (the denominator).

These drawbacks led us to adopt another method for evaluating model quality. We calculated log likelihood ratios to express model improvement over a “naïve model.” In our case the naïve model is simply that our best estimate of the probability of encountering a species in the study area is the average frequency of occurrence of that species in our data. The ratio of the likelihood of the observed values ($\mathbf{y} = y_1, y_2, \dots, y_n$) under the posterior model (M_1) to the likelihood of the result under the naïve model (M_2) is given by:

$$B_{12} = \frac{p(\mathbf{y}|M_1)}{p(\mathbf{y}|M_2)}$$

where

$$p(\mathbf{y}|M) = \prod_{i=1}^n \hat{y}_i^{y_i} (1 - \hat{y}_i)^{1-y_i}$$

and \hat{y}_i corresponds to the fitted values for the likelihood of occurrence under each model, $M_j, j=1,2$. This last formula is the joint probability function for the n observations of y_i , where each observation is an ordinary Bernoulli random variable. For a clear introduction to this in the context of logistic regression, see Neter et al. (1996, p. 573-574). When this formula is applied to the fitted model, the y -hats are the probability estimates from the model. When applied to the naïve model, the y -hats are all the same: simply the overall proportion of 1’s in the response variable.

Formal hypothesis testing with log-likelihood ratios requires that the parameters for one model be nested within the other and incorporates the difference in degrees of freedom between the two models. $\text{Log}_{10}B$ is applied here, however, as a descriptive statistic in the sense of “weight of evidence,” similar to a Bayes factor (Kass & Raftery 1995), rather than a formal hypothesis test. $\text{Log}B$ differs from a Bayes factor in that a Bayes factor is based on the marginal distribution of \mathbf{y} given the prior model (the naïve model in this case), while $\text{log}B$ is a simple log likelihood ratio for the two models, inverted so that as the weight of evidence increases, $\text{log}B$ increases. Values of $\text{log}B$ reported here from NPMR models are based on cross-validated estimates from the M_1 using a leave-one-out strategy. $\text{Log}B$ can be interpreted as the ratio of the likelihood of cross-validated estimates from the fitted model to estimates from the naïve model expressed in powers of ten. $\text{Log}B$ is negative when cross-validated estimates from the fitted model are worse than the naïve model. The same rationale can be applied to the difference between $\text{log}B$ values calculated for each of two competing models of interest. Because $\text{log}B$ is unbounded, it can be quite large when a strong relationship is modeled with a very large data set. The average contribution of a sample unit to $\text{log}B$, $10^{(\text{log}B)/n}$, can be used to describe the strength of relationship, independent of sample size.

Drop in deviance is a statistic closely related to $\text{log}B$, but often used in connection with model evaluation in logistic regression. Similar to a likelihood ratio, the drop in deviance between two models expresses the improvement of one model over the other. Most germane to this discussion is the deviance comparing a particular model against a model with no parameters (i.e. a null model with no predictors):

$$\text{drop in deviance} = 2 [\log \text{likelihood}(\text{model}) - \log \text{likelihood}(\text{null model})]$$

Drop in deviance is linearly related to $\text{log}B$ comparing a model with the naïve model:

$$\text{drop in deviance} = 2 \ln (B_{12})$$

Some software for GLM will report a null deviance and a residual deviance. The drop in deviance is sometimes referred to as the “deviance” for short. One can calculate a $\log B$ from these statistics, assuming the special case of the naive model as described above. In this case,

$$\log B = \frac{\text{null deviance} - \text{residual deviance}}{4.60517}$$

In some accounts the drop in deviance is represented by χ^2 (many authors) or G (Sokal & Rohlf 1995, p. 691) Thus, the difference between “drop in deviance” and $\log B$ in this context is thus more a matter of conceptual alliance than substance. If you have a $\log_{10} B$ statistic but wish to report drop in deviance or χ^2 , simply calculate

$$\begin{aligned}\chi^2 &= 2 (2.302585) \log_{10}(B_{12}) \\ &= 4.605170 [\log_{10}(B_{12})]\end{aligned}$$

Note that the likelihood ratio λ , calculated for a likelihood ratio test, is the inverse of B_{12} as defined above, in that the more complex model is in the denominator. In that case λ decreases as the fitted model improves. The likelihood ratio is converted to a test statistic as $\chi^2 = -2 \ln \lambda$.

One can test the statistical significance of the likelihood ratio using a chi-square distribution with one degree of freedom. Conceptually this requires that the parameter space of one model be considered a subset of, or nested within, the other model. This seems reasonable when comparing any model to the naive model, but questionable when comparing one trial model to another.

To help guard against overfitting and provide a built-in method of cross-validation, we exclude species occurrence at the target point i from the estimation of \hat{y}_i . Conceptually this is a “leave-one-out” strategy (Fielding & Bell 1997) similar to a jackknife estimator: we sacrifice some information to obtain a better estimate of model quality, an estimate with error that is more comparable to the application phase of modeling. Inclusion of the target point would otherwise lead to overfitting the model, because as the window size becomes smaller, point i has a larger influence on \hat{y}_i , and that influence is always in the “correct” direction. In other words, the circularity in logic results in \hat{y}_i with the target included always being as close or closer to y_i than \hat{y}_i with the target excluded.

A special problem occurs with the “hard-edged” window of the uniform kernel. If the window is so narrow that it includes a single point, \hat{y}_i cannot be calculated with target excluded, but $\hat{y}_i = y_i$ (always) with target included. Clearly, including the target point favors narrow windows and leads to overfitting. Excluding the target point eliminates that problem and makes the error rate of the evaluation phase more comparable to the application phase.

Excluding the target point does, however, create a computational problem when ($\hat{y}_i = 1$ and $y_i = 0$) or ($\hat{y}_i = 0$ and $y_i = 1$). These cases are undefined when applied to the likelihood equation above. In these cases, we know that the estimate \hat{y}_i can be improved by including the target point, and that this contributes useful indication of a failure of the model, since the target point disagrees with the remaining points in the window. In this case, including the target point improves our estimate of \hat{y}_i without contributing toward overfitting the model, so we choose to retain the target point. For example, if the window contains two points, $y=[0,1]$, and the first point is the target point, then $\hat{y}_1 = 0.5$ with target included, but with the target point excluded the data point cannot be used in the likelihood equation because $\hat{y}_1 = 1$ and $y_1 = 0$.

The problem described above does not occur with a Gaussian kernel function, because all non-target points contribute at least a minute amount to the estimate. This means that as long as both presences and absences occur in the data, the estimate of \hat{y}_i will never exactly equal zero or one.

Quantitative response data

When the response variable is declared as quantitative, model quality is evaluated in terms of the size of the cross-validated residual sum of squares (RSS) in relationship to the total sum of squares (TSS). We call this a “cross R^2 ” (Antoine & McCune 2004) as implemented in HyperNiche, because the calculation incorporates a cross-validation procedure. The cross R^2 differs from the traditional R^2 because data point i is excluded from the basis for estimating \hat{y}_i . Consequently, with a weak model, it is not uncommon for $RSS > TSS$ and thus cross R^2 becomes negative.

$$xR^2 = \text{cross } R^2 = 1 - \frac{RSS}{TSS} = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2}$$

This approach is essentially the same as the G -value of Agterberg (1984), Gotway et al. (1996), and Guisan and Zimmerman (2000). This same statistic can also be applied to the correspondence between the predicted and observed values for an independent data set, the predictions based on the calibration data.

Sensitivity Analysis

“Sensitivity analysis” has various meanings. Here we consider sensitivity analysis to be an evaluation of the relative importance of particular parameters within the model. Use it to evaluate the importance of individual quantitative predictors in NPMR models. This is particularly important in this context, because with NPMR we have no fixed coefficients or slopes that we can compare in size. The tolerances are related to the importance of variables, but in different ways for local mean and local linear models.

With local mean models, tolerance is inversely related to the importance of a variable. With local linear models, this is not so, because a large tolerance can be obtained in either of two conditions, a strong globally linear effect, or a weak effect. On the other hand a narrow tolerance in a locally linear model implies a strong nonlinear global relationship.

A general method of evaluating the importance of individual variables is to analyze the sensitivity to changes in the variables. One way to do this is to nudge up and down observed values for individual variables, and measure the resulting change in the estimate for that point. By accumulating those sensitivities across all of the data points, or across a large sample of data points, we can evaluate the sensitivity of the model to each predictor. The greater the sensitivity, the more influence that variable has in the model.

The change in the response can be measured as a fraction of the observed range of the response variable. Scaling the differences in response and differences in predictors to their respective ranges allows a sensitivity measure that is a ratio, independent of the units of the variables.

The change in the response can be measured as a fraction of the observed range of the response variable. Scaling the differences in response and differences in predictors to their respective ranges allows a sensitivity measure that is a ratio, independent of the units of the variables.

The general concept is:

$$\text{Sensitivity} = \frac{\text{mean difference in response} / \text{range in response}}{\text{difference in predictor} / \text{range in predictor}}$$

The numerator, the scaled difference in response, is:

$$\text{scaled difference in response} = \frac{\sum_{i=1}^n |\hat{y}_{i+} - \hat{y}_i| + \sum_{i=1}^n |\hat{y}_{i-} - \hat{y}_i|}{2n} \div |y_{\max} - y_{\min}|$$

while the denominator, the scaled difference in the predictor, is Δ . This is the amount by which we choose to nudge the predictor. Combining these yields

$$\text{Sensitivity1} = \frac{\left(\sum_{i=1}^n |\hat{y}_{i+} - \hat{y}_i| + \sum_{i=1}^n |\hat{y}_{i-} - \hat{y}_i| \right)}{2n |y_{\max} - y_{\min}| \Delta}$$

or

$$\text{Sensitivity2} = \frac{1}{|y_{\max} - y_{\min}| \Delta} \sqrt{\frac{\sum_{i=1}^n (\hat{y}_{i+} - \hat{y}_i)^2 + \sum_{i=1}^n (\hat{y}_{i-} - \hat{y}_i)^2}{2n}}$$

where

\hat{y}_{i+} = estimate of the response variable for case i , having increased the predictor by an arbitrarily small value Δ (say 0.1 of the range of the predictor).

\hat{y}_{i-} = estimate of the response variable for case i , having decreased the predictor by an arbitrarily small value Δ (say 0.1 of the range of the predictor).

Δ = A small difference applied to a predictor, expressed as a constant proportion of the range of a predictor

The first term in the equation scales the deviation in the response variable to a proportion of the range in the response variable. Thus, a sensitivity of 1.0 would mean that a 10% change in the predictor would, on average, produce a 10% change in the response variable.

The second term in the equation is either the mean absolute deviation or the square root of the average squared deviation in response resulting from nudging the predictor by Δ .

HyperNiche nudges the predictors one at a time by + or -5% of the range of the predictor. Sensitivity 1 gives less weight to occasional large differences, while Sensitivity 2 will emphasize large differences more, because of squaring the differences. Unless you have a specific reason for choosing the second measure, we recommend using Sensitivity 1. Its interpretation is straightforward. With sensitivity formula 1, a value of 1.0 means that on average, nudging a predictor results in a change in response of equal magnitude. Sensitivity =0.5 means that the response is half the magnitude of the change in the predictor. Sensitivity = 0.0 means that nudging a predictor has no detectable effect on the response.

Model selection

Defining the best model

We seek the best NPMR model by selecting the set of predictor variables and choosing the tolerance for each of those variables. (Choosing a tolerance is the same as selecting a bandwidth, smoothing parameter in the statistical literature.) With more than a few predictors, the number of combinations of predictors and their tolerances are astronomical. Therefore we must use a guided search for the best model. Three criteria determine the best model:

- A measure of model fit (discussed above)
- One or more rules for parsimony (how much improvement in fit that we demand for adding a variable)
- A setting for controlling flexibility, based on the average amount of data used to estimate the response for an individual point (the minimum average neighborhood size, N^*).

Once these three criteria are defined, model selection requires a trial-and-error search through the possible models. Search methods can be broadly categorized into exhaustive, stepwise, and tuning (Table 2). Many different algorithms are possible for stepwise searches and model tuning. The methods used in HyperNiche version 1.0 are described in a section below.

The methods for controlling overfitting differ between NPMR and the generalized linear modeling (GLMs). The most popular overfitting controls for GLMs are the AIC (Akaike Information Criterion) and the BIC (Bayesian Information Criterion) for model selection. The AIC and BIC depend on the number of parameters in a model. Because NPMR models do not have explicit parameters as such, these are not applicable to NPMR models. Instead, use the controls on overfitting provided in HyperNiche (minimum average neighborhood size, minimum data:predictor ratio, and the improvement criterion), as explained below.

Nonparametric regression models sometimes use an AIC based on the "effective number of parameters" (Hastie et al. 2001, p. 205; Hurvich et al. 1998). This penalizes a measure of fit by the trace of the smoothing matrix – essentially how much each data point contributes to estimating itself, summed across all data points. Because NPMR in HyperNiche always uses leave-one-out cross validation in the model fitting phase, the trace of the smoothing matrix is always zero, corresponding to zero parameters for the AIC. Thus, NPMR in HyperNiche is already penalizing the fit through cross validation, and the error rate of the training data set is expected to approximate the error rate in a validation data set. In other words, the training error rate approximates the prediction (extra-sample) error rate.

Table 2. Search methods for the best model.

| Search Method | Definition |
|--------------------------|---|
| Free search – exhaustive | All combinations of variables and tolerances are evaluated, given a specified step size for trial tolerances. All models are evaluated if the number of possibilities is fairly small (e.g., < 10,000). |
| Free search - stepwise | Variables are added or deleted incrementally and tolerances are adjusted incrementally according to a specific search algorithm |
| Tuning | Variables are held constant and tolerances are adjusted incrementally |

Controlling Flexibility and Parsimony

Effective modeling with NPMR requires attention to flexibility and parsimony. These are particularly important with small data sets, or clumped sampling from the predictor space, or a large number of predictors compared to the sample size. The built-in online help system of HyperNiche provides more detail and examples than are provided here.

Flexibility can be controlled by both the leave-one-out crossvalidation and by setting a minimum average neighborhood size (N^*) for a model (e.g., for a small data set, 1 = very flexible; 3 = flexible; 10 = stiff). The larger the sample size, the larger the N^* needed to achieve a given degree of stiffness. The smaller the sample size, the more important the effect of the leave-one-out crossvalidation. A reasonable starting value for the minimum average neighborhood size for most data sets is about 5% of the sample size, i.e. $N^* \geq 0.05(n)$. Choose stiffer curves with very small data sets or clumped data distribution along important habitat variables. More flexibility is allowable with large high-density data sets.

Parsimony in number of predictors can be controlled by setting an improvement criterion, expressed as a percentage improvement in model fit when a new variable is added. One can also set a minimum data:predictor ratio. For quantitative responses, the data:predictor ratio is the number of sample units divided by the number of predictors in the model. For binary responses, the data:predictor ratio is the number of observations in the least represented category (presences or absences) divided by the number of predictors in the model. Harrell et al. (1996) suggested a minimum value of 10 for binary data.

With a large sample size, as the number of predictors in a model increases, the fit tends to flatten out, while the average neighborhood size declines (Fig. 12). With a small data set, fit will sooner or later decline as variables are added. This decline occurs because the cross-validation penalty increases as predictors are added. With HyperNiche you will normally see only the beginning of this decline, because once the decline begins, the stepwise search ends.

The degree of continuity of an estimated response surface depends on how little data you require to estimate a point on the surface. This is controlled by setting a minimum neighborhood size for individual points (n_i^*). For a given model, missing portions of the response surface are minimized by setting a small minimum n_i^* . To see only the well-supported portions of a response surface, set a large minimum n_i^* .

If a data point has a neighborhood size (n_i^*) smaller than the user-defined minimum, then that data point is omitted from the calculations and a missing value indicator is assigned to that point. This criterion is called the “minimum neighborhood size required for estimate” by HyperNiche.

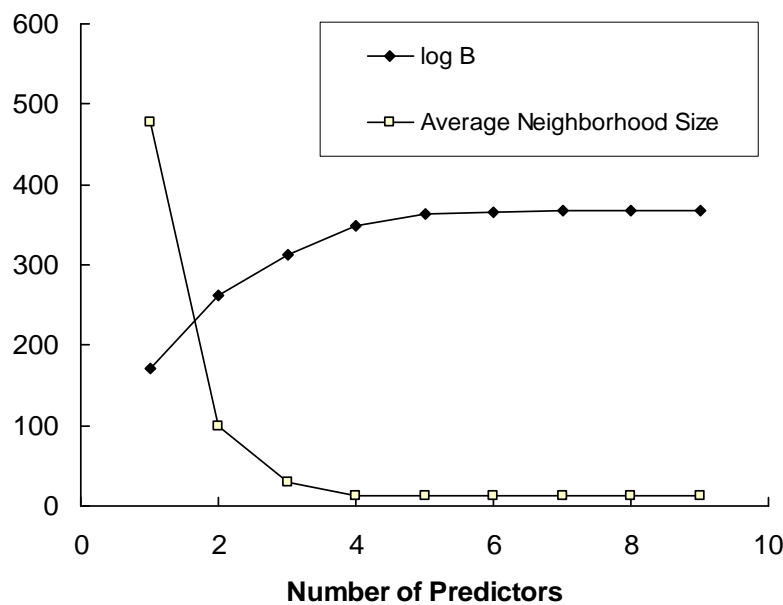


Figure 12. Typical dependence of model fit (in this case measured by $\log B$) and average neighborhood size (N^*) on the number of predictors in the model. In this case the data set was very large ($n = 2500$) and relationships fairly strong, yielding large $\log B$ values. Two and three-predictor models were chosen for further study.

The next example (Table 3) shows the summary table produced by HyperNiche for a binary response variable. This is the same kind of information graphed in Figure 12, but for a different data set, this time with $n = 72$. Use this table to help you decide on the appropriate number of predictors. The change in $\log B$ is itself a $\log B$ comparing the model to the next lower dimensional model. In this case, the second variable contributes some predictive ability, as shown by the “ $\log B$ change” = 0.59. On Kass and Raftery’s scale, this would fall in the “substantial” range, though clearly the second variable is much less useful than the first.

Table 3. Summary table showing the best model for a series of increasing number of predictors. Stratum is a categorical variable, therefore its tolerance is zero. “NumVars” is the number of predictor variables.

| DEPENDENCE OF FINAL MODEL ON NUMBER OF PREDICTORS FOR RESPONSE VARIABLE Lobore | | | | | | | |
|--|---------|-------|-------------|------|--------------|--------------|--------------|
| Response | NumVars | LogB | LogB change | N* | Tol-Variable | Tol-Variable | |
| Lobore | 1 | 3.694 | 3.694 | 23.2 | 0.00-Stratum | | |
| Lobore | 2 | 4.282 | 0.589 | 11.4 | 0.00-Stratum | 0.44-LogDia | |
| Lobore | 3 | 4.277 | -0.005 | 11.4 | 0.00-Stratum | 0.44-LogDia | 48.45-Height |

Expressing the improvement of a higher dimensional model vs. a lower dimensional model is simple. Because we are using log values, you can calculate the $\log B$ for contrasting two non-naive models just by subtracting their $\log B$ s vs. the naive model. So, for example, if $B_{2,0} = 50$ for the 2-predictor model (model 2) vs. the naive model (model 0) and $B_{3,0} = 100$ for the 3-predictor model (model 3) vs. the naive model, then $\log B_{2,0} = 1.7$ and $\log B_{3,0} = 2.0$. To get $\log B_{3,2}$ for 3-predictor vs 2-predictor models, just subtract the two: $\log B_{3,2} = \log B_{3,0} - \log B_{2,0} = 2.0 - 1.7 = 0.3$.

Can the difference in $\log B$ between two models be used as an automatic cutoff value for when to stop trying to add variables? Yes, one could use either a proportionate change or fixed change in $\log B$ as an automatic cutoff value. For example, one could declare that the increment to $\log B$ must beat 1.0 to make adding the variable worthwhile. This is fine when working with a data set of a particular size, but it doesn’t work well as a prescription across data sets that vary greatly in size. With very large data sets (e.g., thousands of sample units), $\log B$ can be huge, and the model will end up with too many variables, because with a huge sample size, even a weak predictor can increase $\log B$ by 1.

Model application

A nonparametric regression model can be applied in many ways (Fig. 13), essentially the same ways that one can apply traditional regression model. A key difference, however, is that estimates from the model always require reference to the original data. In this sense, the calibration data set is an essential part of the model. Those data, combined with a list of predictors and their tolerances allow estimates for new data points (Fig. 13). Such estimates are necessary for many GIS applications or any other application of the model to new data.

The fact that the data are part of the model means that as new data points become available, the model can immediately take them into account. New estimates can be made based directly on the revised data set, or the fit of the model can be refined before calculating estimates for particular cases.

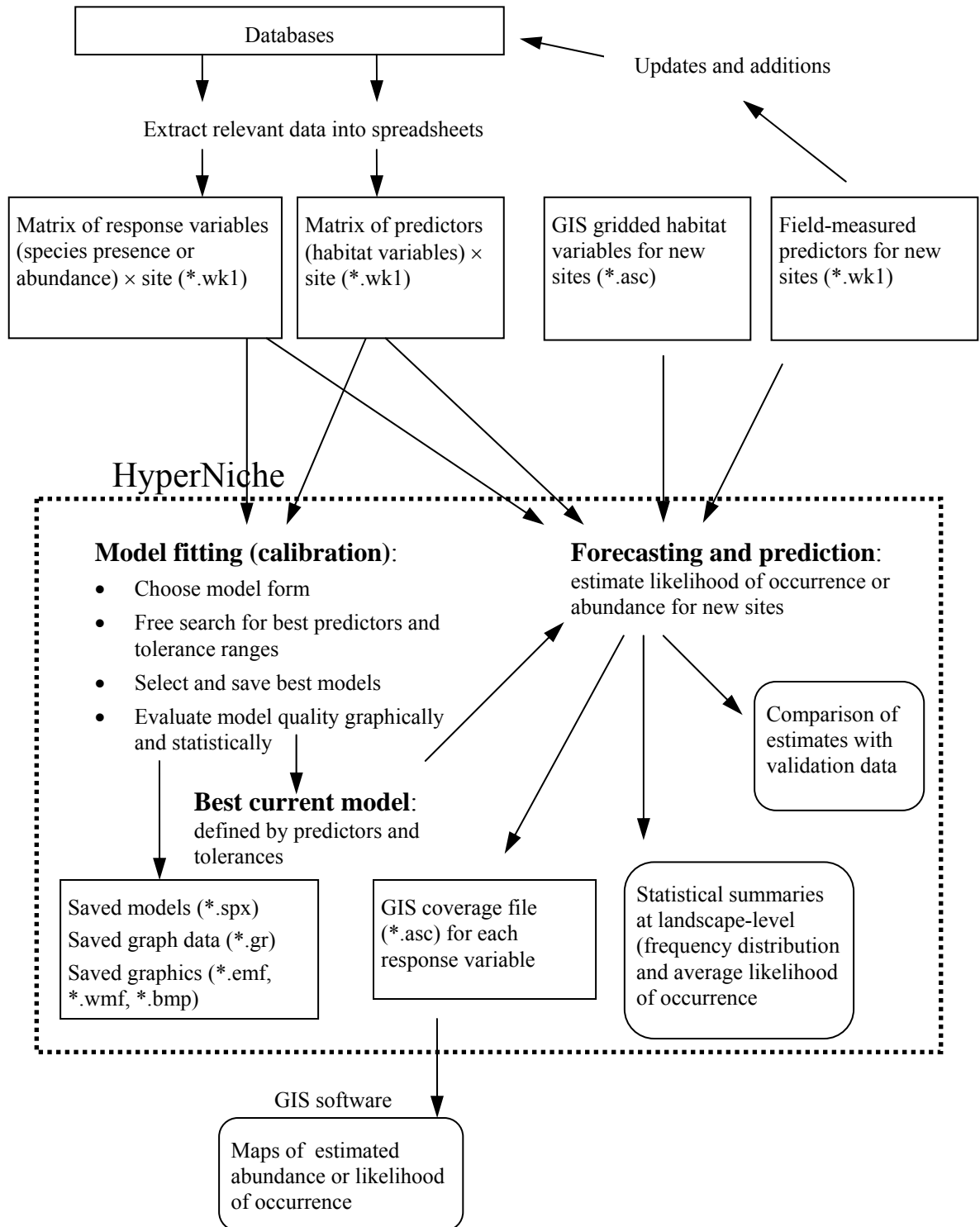


Figure 13. Estimating likelihood of occurrence or abundance of species with the HyperNiche. Square-cornered boxes are files; rounded-corner boxes are products. The dotted box includes components of the program HyperNiche.

What to report for NPMR

What you report will depend on the audience and purpose of your studies. The list suggests some important options in NPMR that should be reported to the technical audience of ecological journals:

Methods

- Statement of the response variable and the pool(s) of predictors.
- Summary of the method. Because NPMR will be unfamiliar to many readers, briefly summarize the method in a phrase or two.
- Local model (local mean or local linear are currently available in HyperNiche).
- Choice of kernel (Gaussian and uniform are currently available in HyperNiche).
- Evaluation of fit with a brief explanation of the measure (xR^2 and $\log B$ are currently available in HyperNiche).

Example of Methods

Berryman and McCune (2006) described their use of NPMR as follows:

NPMR uses a local multiplicative smoothing function with leave-one-out cross-validation to estimate the response variable. We used a Gaussian weighting function with a local mean estimator in a forward stepwise regression of biomass against the predictors, then expressed fit as a cross-validated R^2 (or xR^2).

The xR^2 differs from the traditional R^2 because each data point is excluded from the basis for the estimate of the response at that point. Consequently, with a weak model, the residual sum of squares can exceed the total sum of squares and thus xR^2 becomes negative. Rather than fitting coefficients in a fixed equation, NPMR fits 'tolerances', the standard deviations used in the Gaussian smoothers.

Results

For important or final models, consider reporting as a minimum:

- How well the model fit (xR^2 or $\log B$)
- Which predictors were selected
- Tolerances for each predictor
- Sensitivity for each predictor

Also consider:

- 2D or 3D plots of the fitted response surface
- Plots of residuals or predicted vs. observed
- For each pool of predictors, how fit changes as the number of predictors is increased (e.g. plot xR^2 vs. number of predictors)
- Slicing high-dimensional response surfaces by specifying median (or other selected values) for some predictors, displaying the response surface on the remaining predictors.

Examples of results

Please see the papers listed in Appendix C.

Example Fits for the Gaussian-Gaussian Problem

The data set

By fitting models to a data set with a known underlying structure, we can see how well different models recover that structure. Consider a data set where species abundance responds perfectly to two ecological factors, Factor1 and Factor2. The key question is, how well can different statistical models represent a realistic (but noiseless) response surface? Because there is no noise in the data, a good model should achieve near perfect prediction.

A response surface was designed that incorporates Gaussian responses to two primary factors. A smooth, noiseless surface was generated by multiplying two Gaussian functions.

$$y = \left(e^{-a_1(x_1-b_1)^2} \right) \left(e^{-a_2(x_2-b_2)^2} \right)$$

The first term is the Gaussian response to Factor 1 (x_1) and the second term is the Gaussian response to Factor 2 (x_2). For simplicity, set parameters, \mathbf{b} , to zero, so that the response peaks at (0,0) and $a_1 = a_2 = 1$ to give a standard deviation of 1 for both curves.

$$y = \left(e^{-(x_1)^2} \right) \left(e^{-(x_2)^2} \right) 100$$

The final scaling constant brings the maximum value of y to 100. This surface was then sampled at random with 200 points and the response at those points calculated with this deterministic equation. The points were drawn from uniform random distributions with $(-1 < x_1 < 4)$ and $(-2 < x_2 < 2)$. Note that these ranges means that we are truncating x_1 at -1, but including most of its positive asymptote. For Factor 2, we are truncating our sample evenly on both sides of the hump, at -2 and + 2. The resulting points on the surface of this function form a hump shape, truncated to different degrees on different sides (Fig. 14A). If either factor is unfavorable, then the species is absent or nearly so.

Now pretend that you do not know how the data set was constructed. Put on your biologist hat and try to build a model of the response variable (biomass, y) in relation to Factor 1 and Factor 2. We will build least-squares regression models and multiplicative (NPMR) models to describe the relationship.

Least squares models

One of the simplest and most naïve models would be a least-squares multiple regression with two terms and a constant:

$$y = b_1x_1 + b_2x_2 + b_0$$

Fitting this model to the data (the 200 sample points on the response surface), we find that the fit is rather poor (adjusted $R^2 = 0.41$; Table 4). The resulting equation fits a tilted plane to the data (Fig. 14B). Only the term for Factor1 differs significantly from zero, because the Gaussian response to Factor 2 is nearly balanced, with the peak response falling in the middle of the range of Factor 2. Including an interaction term,

$$y = b_1x_1 + b_2x_2 + b_3x_1x_2 + b_0$$

results in no improvement (adj. $R^2 = 0.41$).

Most analysts would examine the residual plot and see that hump-shaped relationships that are not being fit. A possible solution to this is to include a quadratic term for each factor. Adding these terms gives a pool of five predictors and the model:

$$y = b_1x_1 + b_2x_2 + b_3x_2^2 + b_4x_1x_2 + b_5x_1x_2^2 + b_0$$

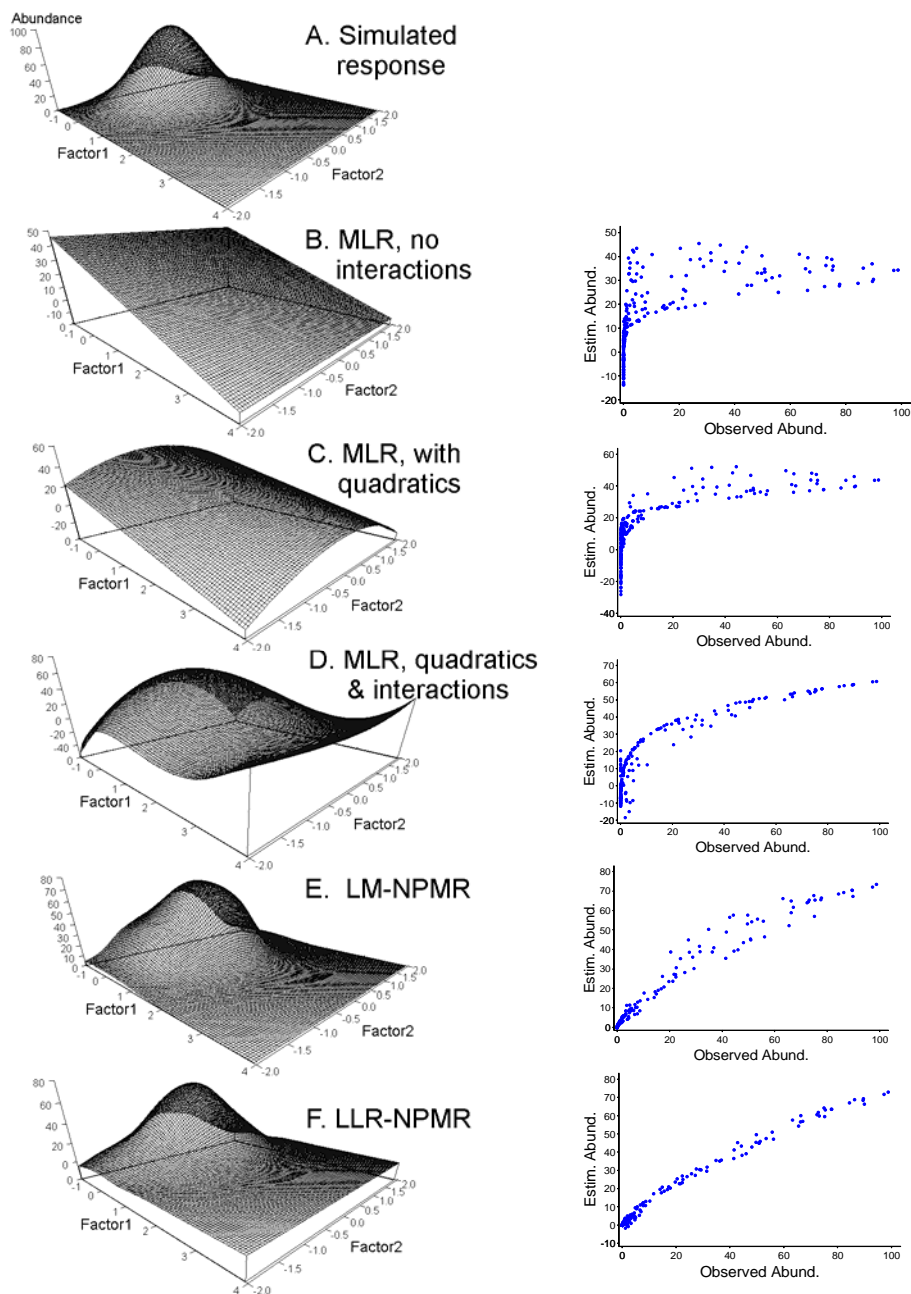


Figure 14. Attempts by various modeling strategies to recover a surface representing Gaussian responses to Factor1 and Factor2. **A.** The simulated response surface that the models try to recover; 200 points were sampled on this surface. **B.** Multiple linear regression (MLR) with two linear terms. **C.** MLR with quadratic terms. **D.** MLR with the best four variables drawn from a pool of 10 variables with quadratic, cubic, and interaction terms of the quadratics with the two factors. **E.** NPMR, local mean and Gaussian weights. **F.** NPMR, local linear model and Gaussian weights. **Right panels:** Comparisons of predicted vs. observed responses.

Using a backwards elimination along with criteria of an R^2 improvement of 0.05 and a $t = 2.0$ for the partial regression coefficients, two variables are retained:

$$y = b_1x_1 + b_2x_2^2 + b_0$$

This model is considerably better (adjusted $R^2 = 0.55$; Table 4), but the surface (Fig. 14C) still does not resemble the actual response surface (Fig. 14A).

Last, we increase the pool of variables to 10 by including cubic terms of the original variables, along with interactions of the quadratic terms with the other predictors. Backwards elimination results in a model with four predictors and adjusted $R^2 = 0.77$ (Table 4). Now our model is:

$$y = b_1x_1^2 + b_2x_2^2 + b_3x_1^2x_2 + b_4x_1x_2^2 + b_0$$

Still the fit is not what we would hope for. Even taking the unparsimonious approach of forcing all ten terms improves the fit to adjusted $R^2 = 0.81$. The resulting surface (Fig. 14D) starts to resemble our known underlying model, but still leaves much to be desired, for example the odd upswing in the surface at high values of Factor1.

NPMR models

Table 4 compares the preceding regressions with results from NPMR, using the same data. The minimum average neighborhood size was set at 10. NPMR easily captured about 95% of the variation in the response variable using the two predictors, Factor1 and Factor2. In contrast, our best least-squares model explained about $\frac{3}{4}$ of the variation in abundance. Note that we did not need to specify interaction terms, because NPMR automatically models interactions.

Table 4. Fitted models for least-squares regression and nonparametric multiplicative regression (NPMR). The R^2 statistics are adjusted R^2 values for least-squares models, and cross-validated xR^2 for NPMR.

| Model type | R^2 |
|--|-------|
| Least-squares regression | |
| 2 factors, no interaction | 0.406 |
| With Factor1 x Factor2 | 0.406 |
| With quadratic terms | 0.554 |
| Best 4 predictors from pool with quadratic, cubic, and interaction terms with quadratics | 0.774 |
| NPMR | |
| Local mean, uniform weights (SpOcc) | 0.949 |
| Local mean, Gaussian weights | 0.930 |
| Local linear, Gaussian weights | 0.945 |

The fitted response surfaces from NPMR (Fig. 14E and 14F) essentially reproduce the original surface. A tight relationship is obtained between predicted and observed values (Fig. 14, right panel). The residuals are relatively well behaved, being fairly small and evenly distributed across the predictors.

The differences among the NPMR models is minor (Tables 4 and 5), compared to the difference between the least squares (MLR) models and the NPMR models. The uniform weights and local mean yielded the highest R^2 , but the 3D plot (not shown) reveals a surface is somewhat bumpy and irregular, and

is probably overfitting slightly. The local linear model performs slightly better than the local mean with Gaussian weights, but at a small cost of producing some estimates that fall outside the range of the original response variable (negative abundances). Considering this fact, the local mean with Gaussian weights seems preferable to the local linear model with this data set.

For this particular data set, a GAM with (poisson family, log link, spline smoother) should readily capture the response surface, because the log link effectively decomposes the two multiplied underlying functions and the smoothing splines capture their shapes.

If, on the other hand, curve shapes do not permeate each dimension of the predictor space, then GAM is likely to perform worse than NPMR (see McCune 2006, App. 3. In other words, GAMs appears to fall short when parallel slices of the response surface along a given predictor have fundamentally different shapes, for example sigmoid on wet sites and hump-shaped on mesic sites. With NPMR, on the other hand, the curve shapes in one part of the multidimensional response surface need not bear any relationship to the shapes in other parts of the response surface.

Table 5. NPMR model characteristics, applied to the simulated response surface combining sigmoid and Gaussian curves. Tolerance is one standard deviation of the Gaussian smoothing function or, in the case of uniform weights, the half-width of the observational window. The column xR^2 is the cross-validated R^2 and N^* is the average neighborhood size (both of these explained below).

| Model type | xR^2 | N^* | Tolerance | |
|---|--------|-------|-----------|---------|
| | | | Factor1 | Factor2 |
| Local mean, uniform weights (SpOcc) | 0.949 | 10.9 | 0.74 | 0.39 |
| Local mean, Gaussian weights (LM-NPMR) | 0.930 | 10.8 | 0.49 | 0.39 |
| Local linear, Gaussian weights (LLR-NPMR) | 0.945 | 10.8 | 0.49 | 0.39 |

Detailed Example of NPMR for a Tiny Data Set

The following example is worked out in detail. The example uses a local mean and a Gaussian kernel. The left hand column gives the algebra for each step, paralleled by the right-hand column which substitutes the numbers of a particular tiny data set.

| Step | Example |
|--|---|
| <p>1. Define a response variable \mathbf{y} and a matrix \mathbf{X} with m predictors. Both matrices have n rows (sample units).</p> <p>Response matrix Predictor matrix</p> $\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & & \\ \vdots & & \ddots & \\ x_{n1} & & & x_{nm} \end{bmatrix}$ | <p>The example has $n = 10$ sample units (rows) and $m = 2$ predictors.</p> $\mathbf{y} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 3 \\ 6 \\ 2 \\ 4 \\ 6 \\ 1 \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} 1 & 5 \\ 3 & 5 \\ 5 & 4 \\ 1 & 3 \\ 3 & 3 \\ 4 & 2 \\ 2 & 1 \\ 3 & 2 \\ 5 & 1 \\ 2 & 4 \end{bmatrix} \quad \bar{y} = 2.4$ |
| <p>2. Set the tolerance, s_j, the standard deviation of the Gaussian weighting function for each predictor. This is the smoothing parameter. Normally this is optimized by an iterative search, but for this example we choose a single value for each predictor.</p> <p style="text-align: center;">tolerance = $s_j, \quad j=1, \dots, m$</p> | <p>$s_1 = 2.6$</p> <p>$s_2 = 1.2$</p> |
| <p>3. Calculate an estimated response for each target point, ($\hat{y}_v, v = 1, n$) as a weighted local mean for each target point, omitting the target point ("leave-one-out" cross-validation).</p> $\hat{y}_v = \frac{\sum_{i=1, i \neq v}^n y_i \left(\prod_{j=1}^m w_{ij} \right)}{\sum_{i=1, i \neq v}^n \left(\prod_{j=1}^m w_{ij} \right)}$ <p>where</p> $w_{ij} = e^{-\frac{1}{2}[(x_{ij} - v_j) / s_j]^2}$ | <p>For \hat{y}_1 the predictors at the first target point are:</p> <p>$v_1 = x_{1,1} = 1 \quad v_2 = x_{1,2} = 5$</p> <p>The estimated response at the first target point is</p> $\hat{y}_1 = \frac{y_2 w_{2,1} w_{2,2} + y_3 w_{3,1} w_{3,2} + \dots + y_{10} w_{10,1} w_{10,2}}{w_{2,1} w_{2,2} + w_{3,1} w_{3,2} + \dots + w_{10,1} w_{10,2}}$ <p>First find the weights in the first term of the numerator:</p> <p>$w_{2,1} = e^{-.5[(x_{2,1} - v_1) / s_1]^2} = e^{-.5[(3-1) / 2.6]^2} = 0.7439$</p> <p>$w_{2,2} = e^{-.5[(x_{2,2} - v_2) / s_2]^2} = e^{-.5[(5-5) / 1.2]^2} = 1$</p> <p>continuing for the other terms...</p> <p>$w_{3,1} = e^{-.5[(5-1) / 2.6]^2} = 0.3062$</p> <p>$w_{3,2} = e^{-.5[(4-5) / 1.2]^2} = 0.7066$</p> <p>...</p> <p>$w_{10,1} = e^{-.5[(2-1) / 2.6]^2} = 0.9287$</p> |

| Step | Example |
|---|--|
| | <p>$w_{10,2} = e^{-.5[(4-5)/1.2]^2} = 0.7066$</p> <p>Then sum the products of the weights and the observations, and divide by the sum of the products of the weights:</p> $\hat{y}_1 = \frac{0 \cdot 0.7439 \cdot 1 + 1 \cdot 0.3062 \cdot 0.7066 + \dots + 1 \cdot 0.9287 \cdot 0.7066}{0.7439 \cdot 1 + 0.3062 \cdot 0.7066 + \dots + 0.9287 \cdot 0.7066}$ <p>Continuing for all of the target points,</p> $\hat{\mathbf{y}} = \begin{bmatrix} 0.93 \\ 1.04 \\ 1.91 \\ 2.56 \\ 2.48 \\ 3.17 \\ 4.41 \\ 3.41 \\ 3.83 \\ 1.49 \end{bmatrix}$ |
| <p>4. The products of the weights can be stored in a $n \times n$ weighting matrix, \mathbf{W}^*: w^*_{ik} is the weight for point i in estimating point k, and $0 \leq w^* \leq 1$:</p> $w^*_{ik} = \prod_{j=1}^m w_{ijk}$ <p>With leave-one-out cross validation the diagonal contains zeros, since a given data point contributes nothing to the estimated response at that point.</p> $\mathbf{W}^* = \begin{bmatrix} 0 & w^*_{12} & \dots & w^*_{1n} \\ w^*_{21} & 0 & & w^*_{2n} \\ \vdots & & \ddots & \vdots \\ w^*_{n1} & w^*_{n2} & \dots & 0 \end{bmatrix}$ <p>Row totals are the sum of the weights applied to each point, the local neighborhood size or local sample size, n^*. Form \mathbf{N}, a diagonal matrix from the row totals of \mathbf{W}^*:</p> $\mathbf{N} = \begin{bmatrix} n_1^* & 0 & \dots & 0 \\ 0 & n_2^* & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \dots & n_n^* \end{bmatrix}$ | $\mathbf{W}^* = \begin{bmatrix} 0 & .744 & .216 & .249 & .185 & .023 & .004 & .033 & .001 & .656 \\ .744 & 0 & .526 & .185 & .249 & .041 & .004 & .044 & .003 & .656 \\ .216 & .526 & 0 & .216 & .526 & .232 & .023 & .185 & .044 & .514 \\ .249 & .185 & .216 & 0 & .744 & .363 & .232 & .526 & .076 & .656 \\ .185 & .249 & .526 & .744 & 0 & .656 & .232 & .707 & .185 & .656 \\ .023 & .041 & .232 & .363 & .656 & 0 & .526 & .929 & .656 & .185 \\ .004 & .004 & .023 & .232 & .232 & .526 & 0 & .656 & .514 & .044 \\ .033 & .044 & .185 & .526 & .707 & .929 & .656 & 0 & .526 & .232 \\ .001 & .003 & .044 & .076 & .185 & .656 & .514 & .526 & 0 & .023 \\ .656 & .656 & .514 & .656 & .656 & .185 & .044 & .232 & .023 & 0 \end{bmatrix}$ $\mathbf{N} = \begin{bmatrix} 2.11 & 0 & \dots & & & & & & & 0 \\ 0 & 2.45 & \dots & & & & & & & 0 \\ \vdots & \vdots & 2.48 & & & & & & & \vdots \\ & & & 3.25 & & & & & & \\ & & & & 4.14 & & & & & \\ & & & & & 3.61 & & & & \\ & & & & & & 2.23 & & & \\ & & & & & & & 3.84 & & \\ & & & & & & & & 2.03 & \\ 0 & 0 & \dots & & & & & & & 3.62 \end{bmatrix}$ |

5. The average neighborhood size is:

$$N^* = \frac{1}{n} \sum_{i=1}^n n_i^*$$

$$N^* = 2.98$$

6. One can express the basic local mean equation above in matrix algebra by first forming a smoothing matrix, \mathbf{U} :

$$\mathbf{U}' = \mathbf{W}^* (\mathbf{N})^{-1}$$

or

$$\mathbf{U}' = \begin{bmatrix} 0 & w_{12}^* & \dots & w_{1n}^* \\ w_{21}^* & 0 & & w_{2n}^* \\ \vdots & & \ddots & \vdots \\ w_{n1}^* & w_{n2}^* & \dots & 0 \end{bmatrix} \begin{bmatrix} 1/n_1^* & 0 & \dots & 0 \\ 0 & 1/n_2^* & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \dots & 1/n_n^* \end{bmatrix}$$

\mathbf{U} is equivalent to dividing each element of \mathbf{W}^* by its row total, the neighborhood size (n_i^*). One row of \mathbf{U} contains the weights for each element of \mathbf{y} , the weights defined by the proximity in \mathbf{X} of each observation to the target point. Note that \mathbf{U} still has zeros in the diagonal but is asymmetric. \mathbf{U} differs from \mathbf{W} in that the row totals of \mathbf{U} are all 1.0, while the row totals of \mathbf{W} will usually vary. Consequently, the weight applied to point i in estimating a response at point j is not the same as the weight applied to j when estimating at point i . This is because the sum of the weights in \mathbf{W} will usually differ for each point (see \mathbf{N} above).

$$\mathbf{U} = \begin{bmatrix} 0 & u_{12} & \dots & u_{1m} \\ u_{21} & 0 & & u_{2m} \\ \vdots & & \ddots & \vdots \\ u_{n1} & u_{n2} & \dots & 0 \end{bmatrix}$$

This is analogous to the “hat” matrix in ordinary least squares regression, because one can calculate the “y-hats” by multiplying the hat matrix (or smoothing matrix) by the observed responses, \mathbf{y} . The estimated responses at the data points are:

$$\hat{\mathbf{y}} = \mathbf{U}\mathbf{y}$$

Calculate the smoothing matrix:

$$\mathbf{U}' = \begin{bmatrix} 0 & .744 & \dots & .656 \\ .744 & 0 & & .656 \\ \vdots & & \ddots & \vdots \\ .656 & .656 & \dots & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{2.11} & 0 & \dots & 0 \\ 0 & \frac{1}{2.45} & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \dots & \frac{1}{3.62} \end{bmatrix}$$

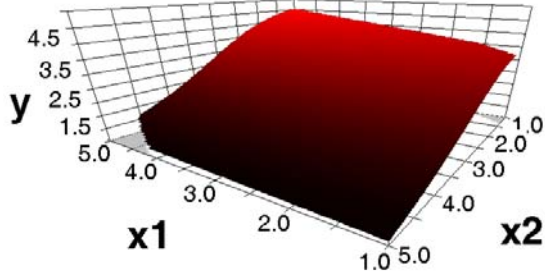
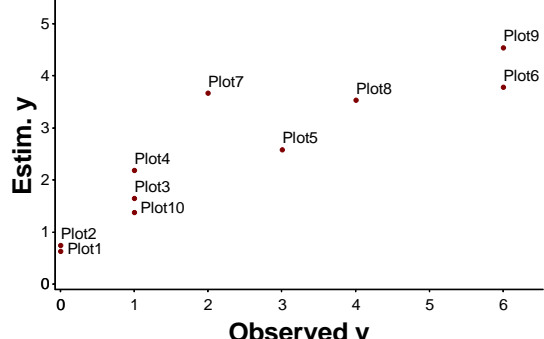
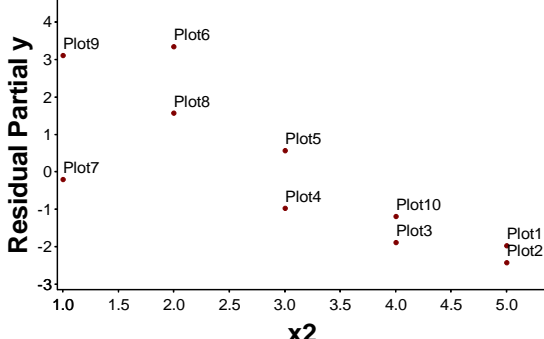
$$\mathbf{U} = \begin{bmatrix} 0 & .352 & .102 & .118 & .088 & .011 & .002 & .015 & .001 & .311 \\ .303 & 0 & .214 & .076 & .102 & .017 & .001 & .018 & .001 & .268 \\ .087 & .212 & 0 & .087 & .212 & .093 & .009 & .075 & .018 & .207 \\ .077 & .057 & .067 & 0 & .229 & .112 & .071 & .162 & .024 & .202 \\ .045 & .060 & .127 & .180 & 0 & .158 & .056 & .171 & .045 & .158 \\ .006 & .011 & .064 & .101 & .182 & 0 & .146 & .257 & .182 & .051 \\ .002 & .002 & .010 & .104 & .104 & .235 & 0 & .294 & .230 & .020 \\ .009 & .011 & .048 & .137 & .184 & .242 & .171 & 0 & .137 & .060 \\ .001 & .001 & .022 & .038 & .091 & .324 & .253 & .259 & 0 & .011 \\ .181 & .181 & .142 & .181 & .181 & .051 & .012 & .064 & .006 & 0 \end{bmatrix}$$

Multiplying the smoothing matrix by the response data:

$$\hat{\mathbf{y}} = \begin{bmatrix} 0 & .352 & .102 & .118 & .088 & .011 & .002 & .015 & .001 & .311 \\ .303 & 0 & .214 & .076 & .102 & .017 & .001 & .018 & .001 & .268 \\ .087 & .212 & 0 & .087 & .212 & .093 & .009 & .075 & .018 & .207 \\ .077 & .057 & .067 & 0 & .229 & .112 & .071 & .162 & .024 & .202 \\ .045 & .060 & .127 & .180 & 0 & .158 & .056 & .171 & .045 & .158 \\ .006 & .011 & .064 & .101 & .182 & 0 & .146 & .257 & .182 & .051 \\ .002 & .002 & .010 & .104 & .104 & .235 & 0 & .294 & .230 & .020 \\ .009 & .011 & .048 & .137 & .184 & .242 & .171 & 0 & .137 & .060 \\ .001 & .001 & .022 & .038 & .091 & .324 & .253 & .259 & 0 & .011 \\ .181 & .181 & .142 & .181 & .181 & .051 & .012 & .064 & .006 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 3 \\ 6 \\ 2 \\ 4 \\ 6 \\ 1 \end{bmatrix}$$

yields:

$$\hat{\mathbf{y}} = \begin{bmatrix} 0.93 \\ 1.04 \\ 1.91 \\ 2.56 \\ 2.48 \\ 3.17 \\ 4.41 \\ 3.41 \\ 3.83 \\ 1.49 \end{bmatrix}$$

| | |
|--|--|
| <p>7. Now evaluate the fit of the estimates to the data. For quantitative data, we use the error (residual) sum of squares (RSS) in proportion to the total sum of squares (TSS):</p> $xR^2 = 1 - \frac{RSS}{TSS}$ $xR^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2}$ | $xR^2 = 1 - \frac{(0-0.93)^2 + (0-1.04)^2 + \dots + (1-1.49)^2}{(0-2.4)^2 + (0-2.4)^2 + \dots + (1-2.4)^2}$ $xR^2 = 1 - \frac{24.6}{46.4}$ $xR^2 = 0.47$ |
| <p>8. Plot a 3D response surface in relation to two predictors, $\hat{y} = f(x_1, x_2, \dots, x_m)$, by estimating a response for selected values of x, normally a grid of combinations of x_1, x_2, etc. For each grid position, estimate the response as a local mean (see equation in Step 3). Do not plot a response if the local neighborhood size is less than a user-set minimum.</p> | <p>Response surface with minimum neighborhood size = 3. Grey cutouts have insufficient data for estimate.</p>  |
| <p>9. Plot predicted (\hat{y}) vs. observed (y) values.</p> |  |
| <p>10. Plot residuals for a partial model, excluding a focal variable.</p> | <p>Residuals for a model without x_1 versus x_2. This shows the relationship of the response to x_2 while controlling for the other variables in the model (in this case only x_1).</p>  |

11. To evaluate the contributions of each predictor to the model, calculate a sensitivity Q to each predictor j . Nudge up and down the observed values for individual predictors, and measure the resulting change in the estimated response for that point

$$Q_j = \frac{\left(\sum_{i=1}^n |\hat{y}_{i+} - \hat{y}_i| + |\hat{y}_{i-} - \hat{y}_i| \right)}{2n |y_{\max} - y_{\min}| \Delta}$$

where

\hat{y}_{i+} and \hat{y}_{i-} are the estimates of the response variable for case i , having increased or decreased, respectively, the value of the predictor by an arbitrarily small proportion, Δ . The estimates \hat{y}_i are calculated in this case without crossvalidation.

$|y_{\max} - y_{\min}|$ = observed range of the response variable

Δ = an arbitrary small proportion of the range of predictor j

The responses are calculated on the basis of the nudged values of the predictors, \mathbf{x} . These are nudged by the proportion Δ of the range in individual columns of \mathbf{x} :

Range = $|x_{\max 1} - x_{\min 1}|$

Amount that x_{ij} is nudged = $|x_{\max 1} - x_{\min 1}| \Delta$

$$|x_{\max 1} - x_{\min 1}| = 4 \quad \Delta = 0.05$$

$$|x_{\max 1} - x_{\min 1}| \Delta = 0.2$$

For the first predictor ($j=1$), nudge each x_1 up or down by 0.3, leaving x_2 unchanged:

$$\text{nudged points} = \begin{bmatrix} 1 + .2 & 5 \\ 1 - .2 & 5 \\ 3 + .2 & 5 \\ 3 - .2 & 5 \\ \vdots & \vdots \\ 2 + .2 & 4 \\ 2 - .2 & 4 \end{bmatrix} = \begin{bmatrix} 1.2 & 5 \\ 0.8 & 5 \\ 3.2 & 5 \\ 2.8 & 5 \\ \vdots & \vdots \\ 2.2 & 4 \\ 1.8 & 4 \end{bmatrix}$$

Estimate y for each of these nudgings, then calculate the difference between the estimate for the nudged value and the estimate for the non-nudged value:

$$\hat{\mathbf{y}}_{\pm} - \hat{\mathbf{y}} = \begin{bmatrix} 0.64 - 0.63 \\ 0.62 - 0.63 \\ 0.75 - 0.74 \\ 0.73 - 0.74 \\ \vdots \\ 1.41 - 1.38 \\ 1.36 - 1.38 \end{bmatrix}$$

Sum the absolute values of the differences and divide by number of nudgings and the fraction of the range in the response variable:

$$Q_1 = \frac{|.64 - .63| + |.62 - .63| + |.75 - .74| + \dots + |1.36 - 1.38|}{2 \cdot 10 \cdot 6 \cdot 0.05}$$

$$Q_1 = 0.115$$

Repeat by nudging the second variable, leaving the first one intact:

$$Q_2 = 0.463$$

The response is about 4 times more sensitive to the second predictor than to the first.

12. Alternatively, to be more consistent with the traditional concept of variance, calculate the sensitivity from the root mean squared differences rather than the absolute differences:

$$Q_j = \frac{1}{|y_{\max} - y_{\min}| \Delta} \sqrt{\frac{\sum_{i=1}^n (\hat{y}_{i+} - \hat{y}_i)^2 + (\hat{y}_{i-} - \hat{y}_i)^2}{2n}}$$

$$Q_1 = \frac{1}{6 \cdot 0.05} \sqrt{\frac{(.65 - .93)^2 + (.61 - .93)^2 + (.76 - 1.04)^2 + \dots + (1.35 - 1.49)^2}{2 \cdot 10}}$$

$$Q_1 = 0.130 \quad Q_2 = 0.507$$

Again, the response is about 4 times more sensitive to the second predictor than to the first.

Comparison of NPMR in HyperNiche to Nonparametric Multiple Regression in S-Plus

The key features of NPMR (nonparametric multiplicative regression) can be understood by comparison to nonparametric multiple regression in the software S-Plus. The following comparison is based on the terminology and logic of Fox (2002).

Fox (2002) gives the nonparametric multiple regression model as:

$$\begin{aligned} y_i &= f(\mathbf{x}'_i) + \varepsilon_i \\ &= f(x_{i1}, x_{i2}, \dots, x_{ik}) + \varepsilon_i \end{aligned}$$

where $\mathbf{x}'_i = (x_1, x_2, \dots, x_k)$ is a vector of k predictors for the i th of n observations and \mathbf{y} is the response. The errors ε_i are often assumed to be independent and normally distributed. The object in NPMR, as in nonparametric regression, is to estimate the response surface rather than to estimate parameters for a model of defined functions relating \mathbf{y} to \mathbf{x} .

1. The first step in estimating the response surface is to define a multivariate neighborhood around a focal point (or target point) $\mathbf{x}'_0 = (x_{01}, x_{02}, \dots, x_{0k})$. The default method in the `loess` function of S-Plus defines the neighborhood by scaled Euclidean distances:

$$D(x_i, x_0) = \sqrt{\sum_{j=1}^k (z_{ij} - z_{0j})^2}$$

where the z_j are predictors standardized to mean = 0 and variance = 1.

2. Next, weights are defined based on the scaled distances:

$$w_i = W\left[\frac{D(x_i, x_0)}{h}\right]$$

where $W()$ is a weighting function and h is the half-width of the neighborhood. In other words, the weight applied to any given point in estimating the value at a target point is based on the distance of the given data point from the target point in the predictor space. Fox (2002) mentions the tricube weighting function, while HyperNiche defaults to a Gaussian weighting function. The value h is also known as the bandwidth or smoothing parameter, or in HyperNiche in the context of species response functions, as the tolerance.

NPMR uses a particular form of this, differing from that in S-Plus in that:

- h is selected for each predictor, so that it is h_j , rather than using a single h for all predictors. This allows us to represent organisms responding strongly to some predictors and weakly to others.
- in HyperNiche h_j is optimized for each predictor in combination with the other predictors, so as to minimize crossvalidation error
- weights for particular data points are always multiplied across predictors in NPMR. If a Gaussian weighting function is chosen, the relationship to the scaled Euclidean distance (above) can be seen:

The weight for a particular point i and predictor j is:

$$w_{ij} = \exp\left\{-\left[\frac{(x_{ij} - x_{0j})}{h_j}\right]^2 / 2\right\}$$

Combining weights multiplicatively for point i across predictors we have:

$$w_i = \prod_j w_{ij}$$

Substituting into this the preceding expression, we get:

$$\begin{aligned} w_i &= \prod_j \exp\left\{-\left[\frac{(x_{ij} - x_{0j})}{h_j}\right]^2 / 2\right\} \\ &= \exp\left\{-\sum_j \left[\frac{(x_{ij} - x_{0j})}{h_j}\right]^2 / 2\right\} \end{aligned}$$

Notice that now the scaled Euclidean distance is weighted separately on each dimension of the predictor space, according to h_j . In essence, the predictor space is stretched or compressed according to the importance of the predictor. The predictor space is stretched in the dimension of a weak predictors, such that the data points appear relatively far (and therefore not influential) in those dimensions.

With NPMR we need not preserve these Euclidean properties or the conversion of the exponential function into an additive function. For example, we may include categorical predictors with $h_j = 0$ or use another nonexponential weighting function.

3. Last, responses are estimated. In nonparametric multiple regression, a weighted polynomial regression is made of y on the predictors. A separate weighted least-squares regression is solved at each target point. While the regression fits a relationship for the neighborhood of the target point, a particular regression is needed to fit the response at each target point. Fox (2002) gave the example of a local linear model:

$$y_i = a + b_1(x_{i1} - x_{01}) + b_2(x_{i2} - x_{02}) + \dots + b_k(x_{ik} - x_{0k}) + e_i$$

Note that the emphasis given to particular predictors by the regression coefficients \mathbf{b} can thus vary throughout the predictor space, according to the strength of their local linear relationships to the response.

Step 3 is the same as above for the local linear form of NPMR, except that the target point is excluded from the fitting process. This leave-one-out cross-validation process makes our error estimates more realistic. With the local mean form of NPMR, the response are estimated as a weighted average rather than a weighted least-squares regression.

The differences between basic nonparametric multiple regression, as outlined in Fox (2002), and NPMR, as implemented in HyperNiche, are summarized in the table below. These concern only the processes of defining the local neighborhood and estimating the response surface. Further differences emerge in the process of variable selection.

Table 6. Comparison of NPMR in HyperNiche with basic nonparametric regression as in Fox (2002).

| Property | Basic nonparametric multiple regression | NPMR in HyperNiche |
|--|---|---|
| weights | single weight based on Euclidean distance to target (isotropic predictor space) | separate weights for each predictor allow us to stretch or compress the predictor space in various dimensions (anisotropic predictor space) |
| information from target point | included in fitting procedure with maximum weight | excluded from fitting procedure |
| bandwidth (smoothing parameter or tolerance) | single bandwidth | separate bandwidth for each predictor |
| bandwidth selection | arbitrary | optimized for each predictor in combination with the other predictors, using minimum cross-validation error as optimization principle |

Example of Overfitting

Overfitting is a potentially serious problem for statistical models, particularly with small sample sizes or when the number of predictors is large, relative to the sample size. The very small data set in Table 7 provides an extreme example of the problems of overfitting, illustrated with standard least-squares multiple regression. Using a cross-validated R^2 clearly reveals this overfitting.

Table 7. A small simulated data set to illustrate the problem of overfitting.

| <i>i</i> | Response variable | Predictor variables | | |
|----------|----------------------|------------------------|------|-----|
| | Y | X | X2 | X3 |
| 1 | 1.0 | 1.0 | 23.0 | 7.0 |
| 2 | 4.5 | 1.5 | 62.0 | 1.0 |
| 3 | 2.5 | 5.0 | 13.0 | 5.0 |
| 4 | 0.5 | 7.0 | 87.0 | 2.0 |
| 5 | 2.0 | 7.5 | 33.0 | 9.0 |

If we regress Y on the Xs, we obtain an R^2 of 0.34. To calculate a cross-validated R^2 , we fit a regression line for the data five times (Fig. 15), each time omitting one of the data points. We then calculate an estimated value for that each point, based on a regression equation that did not include that particular point. Subtracting the observed values from those estimates, we can calculate a leave-one-out residual sum of squares (Table 8). Because we have seriously overfit this model, the cross R^2 is far lower than the regular R^2 , indicating that extreme overfitting.

Table 8. Residual sum of squares (SSR), total sum of squares (SST) and R^2 calculated in the usual way compared with values from a leave-one-out crossvalidation. The large drop from the usual R^2 to the cross-validated R^2 reveals that the data were grossly overfit.

| | Usual statistic | Leave-one-out cross-validated |
|-------|--------------------|----------------------------------|
| SSR | 6.4 | 335.3 |
| SST | 9.7 | 9.7 |
| R^2 | 0.34 | -33.6 |

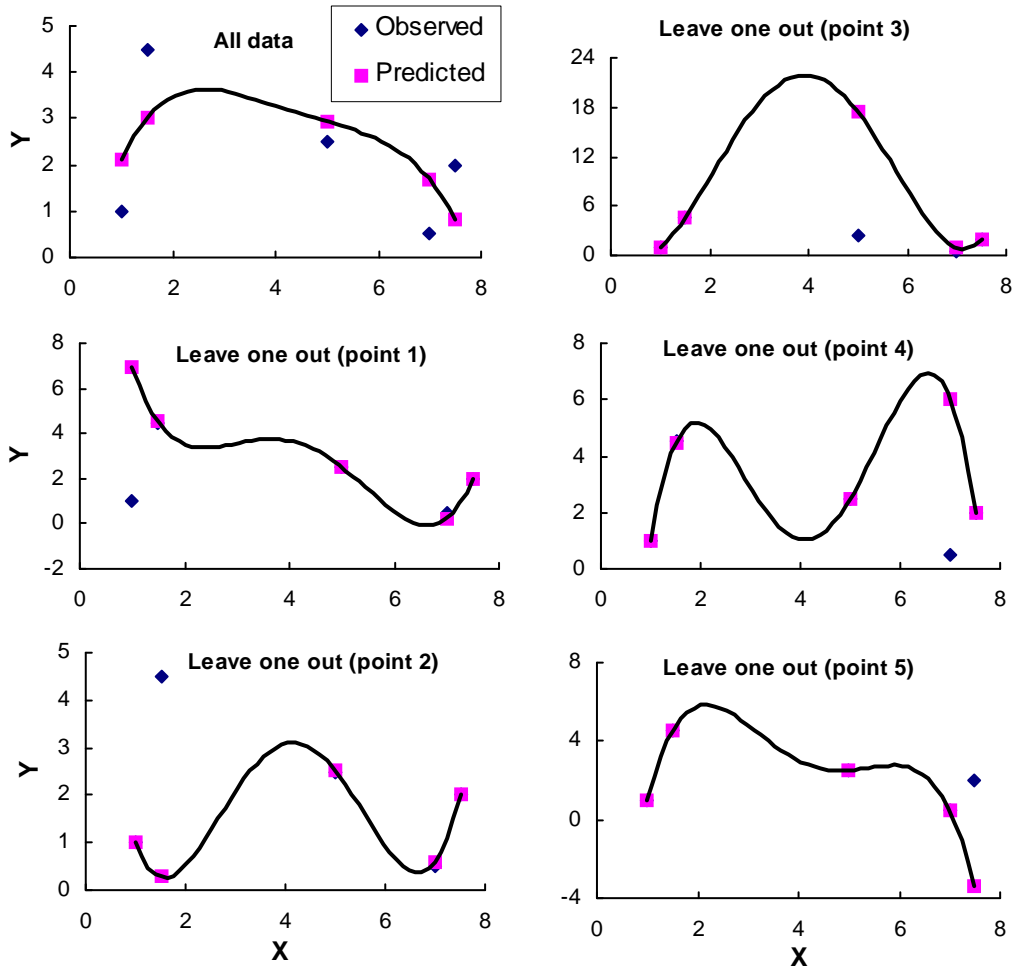


Figure 15. Example of overfitting revealed with leave-one-out cross validation. The upper left panel shows the least-squares fit to $y = f(X, X_2, X_3)$ for the whole data set ($R^2 = 0.34$, $n = 5$). The remaining panels show the curve fit with each point left out ($n = 4$ for each of the “leave one out” panels). The residuals are the vertical distances between the curves and the diamonds. Based on these residuals, the cross-validated $R^2 = -33.6$.

Appendix A – Notation

\mathbf{b} = vector of regression coefficients with $j = 0$ to m elements and b_0 is the intercept.

B_{kl} = likelihood ratio comparing fitted model k to naive model l .

i = array index for n sample units (sites)

j = array index for m predictor variables

k = array index for miscellaneous arrays

l = array index for miscellaneous arrays

m = number of predictor variables in the model

M_k = model k to be compared to another model with a likelihood ratio.

n = number of sample units (sites)

n_i^* = neighborhood size, the amount of data bearing on the estimate of the response variable at point i , calculated as the sum of the weights applied to that particular point.

N^* = average neighborhood size across all data points.

q = number of predictor variables not in the model

s_j = standard deviation of the Gaussian weighting function for predictor variable j , applied to a given predictor such that the full range of observed values for that variable falls over six standard deviations. Also known as the smoothing parameter or bandwidth.

\mathbf{v} = vector specifying the habitat (or other predictors) at the target point, this vector being row vector of $j = 1$ to m columns (variables). This vector usually represents the position of the target point in a space defined by the $j = 1, 2, \dots, m$ predictors.

w_{ij}^* = weight applied to point i for predictor j . The asterisk indicates that it is a univariate weight, as opposed to a weight from the matrix \mathbf{W} .

\mathbf{W} = $n \times n$ diagonal matrix with each diagonal element being a product of weights from each predictor variable. For sample unit i , the diagonal element is

$$w_{ii} = \prod_{j=1}^m w_{ij}^*$$

\mathbf{X} = matrix of predictors (habitat or environmental variables) with $i = 1$ to n rows (sample units) and $j = 1$ to m columns (variables).

\mathbf{y} = vector of observed presence-absence, abundance, or other response variable. This is a column vector of $i = 1$ to n rows (sample units).

\hat{y}_v = fitted value or estimated probability of occurrence of species at target point v . If the vector \mathbf{y} contains presence-absence data then \hat{y}_v is a likelihood of occurrence. If \mathbf{y} contains a measure of abundance then \hat{y}_v is an estimate of abundance.

\mathbf{Z} = design matrix of predictors (habitat or environmental variables) with $i = 1$ to n rows (sample units) and $j = 1$ to $m+1$ columns (variables). The first column contains 1s. The i th row has the elements:

$$[1 \quad (x_{1i} - v_1) \quad (x_{2i} - v_2) \quad \dots \quad (x_{ji} - v_j)]$$

Appendix B – Method for Free Search

This section describes how the program HyperNiche searches for the best combination of predictor variables, as part of its “Free search” option. The “best” model depends on fit, parsimony, and minimum average neighborhood size, as discussed above.

One of two methods is used, either an **exhaustive search** of all possible models or a **stepwise search** using an algorithm for a guided search. The computer automatically uses the exhaustive search if the number of possible models is less than a certain number, set at 100,000 in HyperNiche version 1.0.

An exhaustive search evaluates all possible combinations of predictors and tolerances. For quantitative variables, tolerances are varied in increments of 5% (or another increment of your choice) of the range of the habitat variable. Categorical variables require an exact match, so for these, tolerance is not varied.

HyperNiche 1.0 defaults to 15 tolerance levels at 5% increments (5, 10, 15 ... 75%) for quantitative predictors, plus a 16th level for “off” (i.e. the variable removed from the calculation). Tolerances broader than 75% are not examined because they are rarely useful. With the default of 5% increments, if there are c categorical predictors available, and q quantitative predictors available, then an exhaustive search requires the evaluation of the following number of models:

$$\text{Number of trial models} = (16^q)(2^c)$$

Clearly, the number of trial models increases very rapidly with the size of the pool of predictors. With even a moderate number of potential predictors, a sensible search through the potential models is needed.

The stepwise search begins by screening all predictors for the best one-variable model. Additional variables are then added stepwise, seeking improvement at each step. However, the algorithm looks both forward and backward, in that with the addition of a given variable, each variable already in the model is considered for removal or adjustment in tolerance.

The steps below are for quantitative predictors. Categorical predictors are handled similarly except that tolerances cannot be adjusted -- the variable is simply included or not.

Step 1

Screen predictors for best single-predictor model by calculating model fit for each tolerance level for each predictor. Select from these the model with the best fit as the current model.

Step 2

Define a set of trial models with $m + 1$ predictors, consisting of variable(s) already in the model plus the possible addition of each additional predictor at a full range of tolerances.

If there are c categorical predictors available, m quantitative predictors in the model, q quantitative predictors not in the model, and if the number of trial tolerances for the new variable is NINCR then:

$$\text{Number of trials in a set} = (3^m) (\text{NINCR})(q)(2^c)$$

Step 3

Repeat #2 but adjusting the tolerance of each previous predictor by adding or subtracting the specified increment. For **Method 1** (thorough backtracking) try excluding each combination of those variables. For **Method 2** (minimal backtracking) only one existing variable at a time is adjusted. Method 1 is more thorough, but it wastes time by evaluating models that have already been evaluated. Method 2 seems relatively fast and effective and is recommended.

Example of free search.— With one previous quantitative predictor in the model (VAR1) and one new quantitative predictor (VAR2) tentatively selected, we try 2 tolerances for VAR1 times NINCR tolerances for VAR2 (Table 9). The existing model includes VAR1 with a tolerance of 10% of its range (“In(10)” in Table 9) while VAR2 is not in the model (“Out”). Assume our increment for tolerance is 5% of the range and we test only tolerances from 5-75%. All of the trial models are listed below. The model resulting in the best fit is selected as the

next model for trials for further improvement. Because there is only one variable already in the model, Methods 1 and 2 are the same.

Table 9. Example of free search with two quantitative variables, showing all possible trial models. “In” means the variable is in the model; “out” means it is out of the model. Tolerances are expressed here as a percentage of the range in the predictor.

| Trial number | VAR1 (tolerance) | VAR2 (tolerance) |
|----------------|------------------|------------------|
| Existing model | In (10) | Out |
| 1 | Out | In (5) |
| 2 | Out | In (10) |
| 3 | Out | In (15) |
| etc. | etc. | etc. |
| 15 | Out | In (75) |
| 16 | In (5) | In (5) |
| 17 | In (5) | In (10) |
| 18 | In (5) | In (15) |
| etc. | etc. | etc. |
| 30 | In (5) | In (75) |
| 31 | In (10) | In (5) |
| 32 | In (10) | In (10) |
| 33 | In (10) | In (15) |
| etc. | etc. | etc. |
| 45 | In (10) | In (75) |
| 46 | In (15) | In (5) |
| 47 | In (15) | In (10) |
| 48 | In (15) | In (15) |
| etc. | etc. | etc. |
| 60 | In (15) | In (75) |

If the existing and new predictors are both categorical, and membership in the categories is strictly applied (tolerance = 0), the models in Table 10 are evaluated. If the existing predictor is quantitative and a categorical predictor is being considered for addition, the models in Table 11 are evaluated.

Table 10. Example of free search with two categorical predictors, showing all possible trial models.

| Trial number | VAR1 | VAR2 |
|----------------|------|------|
| Existing model | In | Out |
| 1 | Out | In |
| 2 | In | In |

Table 11. Example of free search with one existing quantitative predictor and a categorical predictor being considered for addition, showing all possible trial models.

| Trial number | Variable 1 (tolerance) | Variable 2 |
|----------------|------------------------|------------|
| Existing model | In (10) | Out |
| 1 | Out | In |
| 2 | In (5) | In |
| 3 | In (10) | In |
| 4 | In (15) | In |

With two predictors already in the model, one quantitative and one categorical, and a third categorical predictor is being considered for addition, the following set of models is evaluated (Table 12).

Table 12. Example of free search with two existing predictors in the model, one quantitative and one categorical. A third categorical predictor is being considered for addition. All trial models are shown.

| Method 1 (thorough backtracking) | | | |
|---|------------------------|------------|------------|
| Trial number | Variable 1 (tolerance) | Variable 2 | Variable 3 |
| Existing model | In (10) | In | Out |
| 1 | Out | In | In |
| 2 | Out | Out | In |
| 3 | In (5) | In | In |
| 4 | In (5) | Out | In |
| 5 | In (10) | In | In |
| 6 | In (10) | Out | In |
| 7 | In (15) | In | In |
| 8 | In (15) | Out | In |

| Method 2 (minimal backtracking) | | | |
|--|------------------------|------------|------------|
| Trial number | Variable 1 (tolerance) | Variable 2 | Variable 3 |
| Existing model | In (10) | In | Out |
| 1 | Out | In | In |
| 2 | In (5) | In | In |
| 3 | In (10) | In | In |
| 4 | In (15) | In | In |
| 5 | In (10) | Out | In |

Step 4

Repeat these trials for each predictor not in the model. With m predictors in the model and q predictors not in the model, the number of trials in this set is:

Method 1: Number of trials in a set = $(4^m)(2^q)$ NINCR

Method 2: Number of trials in a set = $(4m + 2q)$ NINCR

From this total, the number of combinations that result from changing settings of variables already in the model are:

Method 1: Number of combinations for variables already included = $(4^m)(2^q)$

Method 2: Number of combinations for variables already included = $(4m)(2q)$

Step 5

Select the best model from this set. The best model must equal or exceed the minimum average neighborhood size.

Step 6

If the best model from this set is not better than the previous model, according to the parsimony criteria, then stop. Otherwise, accept the best model from this set as the current model.

Step 7

If all of the following are true, go to step 2.

1. the user-defined maximum number of trials has not been exceeded
2. the user-defined Data:Predictor ratio is still exceeded
3. perfect prediction in the cross validation has not been achieved

If any of the above are false, then stop searching for a better model.

Appendix C – Example Publications Using NPMR and HyperNiche

NPMR is relatively new, but the number of papers published with the method has increased rapidly. The following list includes both one-predictor applications (nonparametric regression, NPR), and multipredictor models using the multiplicative weighting functions (nonparametric multiplicative regression, NPMR).

| Reference | Response data type | Notes |
|--|---|--|
| Antoine & McCune (2004) http://dx.doi.org/10.1639/0007-2745(2004)107[0163:CFAREN]2.0.CO;2 | quantitative (growth rates and abundance classes) | Local mean NPMR, Gaussian weights, 1-predictor models, small sample size |
| Berryman & McCune (2006) http://dx.doi.org/10.1658/1100-9233(2006)17[157:EEMBFT]2.0.CO;2 | quantitative (lichen biomass) | Local mean NPMR, Gaussian weights used to relate lichen biomass to stand structure and topography. Based on the response surfaces observed with NPMR, they chose final models of three types: NPMR, nonlinear regression, and multiple linear regression. |
| Binder & Ellis (2008) http://dx.doi.org/10.1017/S0024282908007275 | binary (species presence and randomly generated pseudo-absences in a regional grid) | Local mean NPMR, Gaussian weights. Modeled responses to pollutant loads and climate variables under various climate change scenarios; randomization tests; evaluated spatial autocorrelation. Results further filtered by applying a ‘habitat mask’ representing declines in a key substrate for the target species. |
| Casazza et al (2007) http://dx.doi.org/10.1111/j.1472-4642.2007.00412.x . | quantitative (number of endemic taxa) | Local mean NPMR, uniform weights (“SpOcc” model). Modeled diversity of endemic plants in relation to glacial limit, substrate type, and thermoclimatic belts. |
| Crabtree & Ellis (2010) | quantitative (canopy height, cover of bare ground) | Local mean NPMR, Gaussian weights. Compared with cross-validated errors with multiple least-squares regression. In this case normalized squared residuals were similar between NPMR and MLS. |
| Cristofolini et al. (2008) http://dx.doi.org/10.1016/j.envpol.2007.06.040 | quantitative (lichen diversity) | Local mean NPMR, uniform weights (“SpOcc” model). Modeled overall lichen diversity and nitrophytic lichen diversity in response to pollutant concentrations, stand characteristics, and other environmental variables. |
| DeBano et al. (2010) doi: 10.1603/EN08270 | quantitative (insects trapped per month) | Local mean NPMR and Gaussian weights, modeling insect pests trapped per month against weather and other environmental variables. Numerous 3D wireframe response surfaces, sensitivity analysis, and randomization tests. Clear 1-page explanation of NPMR. |
| Derr et al. (2007) http://dx.doi.org/10.1639/0007- | quantitative (species richness in relation to geography and community) | Local linear NPMR, Gaussian weights. Compared fit of species richness to four different sets of predictors: |

| | | |
|---|--|--|
| 2745(2007)110[521:EMCIPC]2.0.CO;2 | ordination scores) | topographic+geographic, vascular plants, the combination of the two preceding sets, and community ordination scores. |
| Duren & Muir (2010) | quantitative (tree age) | Local mean NPMR, Gaussian weights. Regressed true age against tree size, growth ring patterns, and community type. Used resulting models to estimate tree age for samples with poor quality counts from stem sections. |
| Ellis & Coppins (2007) http://dx.doi.org/10.1658/1100-9233(2007)18[725:CCAHSI]2.0.CO;2 | quantitative (species richness) | Local mean NPMR, Gaussian weights. Stepwise selection of predictors representing climate and forest structure; randomization test. Predictors were selected from a pool of 15 variables and evaluated with a randomization test. Models were used to generate predictions based on future climate scenarios. |
| Ellis et al. (2007a) http://dx.doi.org/10.1016/j.biocon.2006.10.036 | binary (species presence) | Local mean NPMR, Gaussian weights. Modeled species presence against climatic predictors. Applied models to climate change scenarios. |
| Ellis et al. (2007b) http://dx.doi.org/10.1016/j.biocon.2007.08.016 | binary (species presence) for 26 species | Local mean NPMR, Gaussian weights. Modeled species presence against climatic predictors; included randomization tests and AUCs. They present NPMR models for many species, depicted them geographically rather than response surfaces in the predictor space. |
| Fenton & Bergeron (2008) http://dx.doi.org/10.1016/j.biocon.2008.03.019 | quantitative (species richness and evenness) | Assessed the relative roles of age and habitat in creating and maintaining species diversity. "Local mean NPMR, Gaussian weights. "...use of multiple overlapping data sets with NPMR and subsequent comparisons permits complex interactions between different variables to be teased out."..." |
| Flitcroft (2008) https://ir.library.oregonstate.edu/dspace/handle/1957/7262 | quantitative (log density of a species) | Log density of juvenile salmon regressed against habitat characteristics, using local mean and Gaussian weights. Used NPMR because of failure of parametric modeling. |
| Gilligan & Muir (2011) http://www.bioone.org/doi/pdf/10.3955/046.085.0206 | quantitative (tree ages) | Local-linear NPMR, Gaussian weights, regressing tree age against plot and tree characteristics. Included a randomization test and application to predict ages of non-cored trees. |
| Giordani (2007) http://dx.doi.org/10.1016/j.envpol.2006.03.030 | quantitative (diversity) | Local mean NPMR, uniform weights ("SpOcc" model). Diversity regressed against pollutants and other environmental variables. |
| Giordani & Incerti (2008) http://dx.doi.org/10.1007/s11258-007-9324-7 | quantitative (species abundance) | Local mean NPMR, uniform weights ("SpOcc" model). Regressed many species against macroclimatic variables. |
| Grundel & Pavlovic (2008) http://dx.doi.org/10.1650/0010-5422(2007)109[734:ROBSDT]2.0 | quantitative (bird species density) | Local mean NPMR, Gaussian weights, modeling the density of many bird species in relationship to numerous habitat factors. This paper gives a lucid explanation of NPMR, |

| | | |
|---|--|--|
| CO;2 | | three dimensional response surfaces, and some nice examples of interacting nonlinear responses. |
| Hosten et al. (2007) http://www.blm.gov/or/districts/medford/files/livestocksumm.pdf | Grazing utilization | Local mean NPMR, Gaussian weights. Modeled the relationship of maximum utilization and average utilization to environmental factors, vegetative descriptors, and management activities |
| Jovan (2003) http://www.treeseearch.fs.fed.us/pubs/25497 | quantitative (species abundance classes) | Local mean NPMR, Gaussian weights, 1- to 3-predictor models |
| Jovan & McCune (2005) http://www.fia.fs.fed.us/lichen/pdfs/Jovan_and_McCune_2005.pdf | Species abundance in relation to scores on NMS ordinations | Local mean NPMR, Gaussian weights. Used optimum value for a species on one axis while fitting the response curve to another axis. In effect this slices a response surface along a particular plane. |
| Jovan & McCune (2006) http://dx.doi.org/10.1007/s11270-006-2814-8 | Nitrophile abundance in relation to elevation. | Local mean NPR, Gaussian weights, 1 predictor. Compared to nonlinear regression and simple linear regression. |
| Kohler (2007) https://ir.library.oregonstate.edu/dspace/handle/1957/4600 | quantitative, log(abundance of species) | Local mean NPR, Gaussian weights. Regressed population size (density of an insect species) against "hemlock woolly adelgid population score" |
| Lintz et al. (2011) http://dx.doi.org/10.1016/j.ecolmodel.2010.10.017 | quantitative and presence-absence, simulated and real data | Developed indices of threshold strength and diagonality in 3D response surfaces; Local mean NPMR with Gaussian weights. Compared the performance of Random Forests, Classification and Regression Trees, and NPMR using a large variety of 3D response surfaces. They found: "The accuracy of each method depends on the threshold strength and diagonality of the original data structure with each method differing in degree of dependence (Fig. 4). The accuracy of most methods decreases as diagonality increases and threshold strength decreases with the exception of NPMR with continuous data... NPMR demonstrates the least variability (seen as quantile bars in Fig. 4) and the greatest accuracy (seen as medians in Fig. 4) compared to the other methods for a given response shape. The sensitivities of modeling methods to shape attributes of data structure arises from features specific to each modeling method, which manifest in visual differences of predicted surfaces for different shapes (Fig. 5). For our subsequent analyses using real ecological data, we choose the most accurate and robust method we test, NPMR." |
| McCune (2006) https://ir.library.oregonstate.edu/dspace/handle/1957/3685 | binary (species presence) and quantitative (species abundance) | Local mean NPMR, Gaussian weights, medium and large sample sizes, simulated and real data, comparison of linear, logistic and NPMR |

| | | |
|---|--|---|
| | | models. |
| McCune (2007) http://oregonstate.edu/~mccuneb/McCune2007JVS-HeatLoad.pdf | quantitative (potential direct incident radiation) | Local linear NPMR, Gaussian weights, with slope, aspect, and latitude as predictors |
| McCune et al. (2003) http://www.cof.orst.edu/cof/fs/research/silv/berger/sberryman/ESA%20paper.pdf | binary (species presence) | Local mean NPMR, uniform weighting function (predates inclusion of Gaussian weights in HyperNiche), 1- and 2-predictor models |
| Miller et al. (2007) http://dx.doi.org/10.1111/j.1365-2427.2007.01850.x | quantitative (community ordination scores, species richness and density of particular functional groups) | Local mean NPMR, Gaussian weights, used to model stream insect communities in relation to longitudinal gradients. Detailed, clear explanation of NPMR, including model specification and sensitivity analysis. Nice exposition of detecting interactions. |
| Minuto et al. (2006) http://dx.doi.org/10.1080/11263500600756348 | quantitative (genetic diversity) | Local mean NPMR, uniform weights; regressed genetic diversity against geography (latitude, longitude, elevation) and population features (number of individuals, occupancy area, occupancy rate). |
| Ponzetti et al. (2007) http://dx.doi.org/10.1639/0007-2745(2007)110[706:BSCIRT]2.0.CO;2 | quantitative (species abundance classes and community ordination scores) | Local linear NPMR, Gaussian weights; regressed many species against ordination scores; also quantitative community ordination scores regressed against disturbance and cheatgrass. |
| Ponadera & Potapova (2007) http://dx.doi.org/10.1016/j.limno.2007.01.004 | quantitative (abundance of diatom species) | Local linear NPMR, Gaussian weights. Regional-scale analysis of diatom species abundance in relation to water chemistry. |
| Potapova & Wintel (2006) http://www.ansp.org/~potapova/pdfs/Geissleria2006.pdf | quantitative (% relative abundance and log-transformed cell densities) | Local linear and local mean NPMR, Gaussian weights. Modeled abundance of three diatom species in relation to water quality variables. Includes a table comparing fits for local linear and local mean models. In general, the fits were slightly higher for local linear models. |
| Reusser & Lee (2008) http://dx.doi.org/10.1093/icesjms/fsn021 | binary (species presence) in benthic estuarine and coastal communities | Local mean NPMR, Gaussian weights, used to model species presence in relation to habitat and geographic variables at two scales. "NPMR generally performs well at both spatial scales and that distributions of non-indigenous species are predicted as well as those of native species." |
| Rood et al. (2010) http://dx.doi.org/10.1007/s00442-010-1758-2 | quantitative (willow cover) | Local mean NPMR with Gaussian weights, regressing willow cover against environmental variables. Illustrations include 2D response curves superimposed on scatterplots. |
| Schroeder et al. (2010) | binary (tree species presence) | Local mean NPMR with Gaussian weights, relating species presence to climate in a large region. Used three measures of model performance ($\log B$, AUC, % improvement) in relation to range size. They also applied an ensemble approach to base maps of probability of occurrence on the top three models. |

| | | |
|---|--|---|
| | | "Evaluation statistics revealed significant and accurate probability of occurrence models were developed for all six species... Across species, ranked evaluation statistics also revealed a pattern of decreasing model accuracy with increasing range size. At plot level, correlations between dominance and probability of occurrence were weakly positive..." |
| Wedderburn et al. (2007) http://dx.doi.org/10.1111/j.1600-0633.2007.00243.x | quantitative (fish species abundance) | Used NPR to relate individual fish species abundance to salinity. |
| Welsh, H. H. & G. R. Hodgson (2010) | binary (herpetofauna species presence) | Local mean NPMR with Gaussian weights, regressing presence of species against stream and riparian habitat characteristics. |
| Welsh, H. H. et al. (2010) | quantitative (fish and amphibian abundance and amphibian richness) | Local mean NPMR with Gaussian weights, regressing presence of species against aquatic and riparian habitat characteristics. |
| Yost (2006) | binary (species presence) | Local mean NPMR with Gaussian weights, regressing presence of selected species against site factors, past management, and stand characteristics. |
| Yost (2008) http://dx.doi.org/10.1016/j.ecolind.2006.12.003 | binary (species presence) | Local mean NPMR with Gaussian weights. "NPMR was compared with logistic regression (LR) by building reduced models from variables selected as best by NPMR and full models from variables identified as significant with a forward stepwise process and further manual testing. LogB was used to select models with the highest predictive capability. NPMR models were less complex and had higher predictive capability than LR for all modeling approaches. Spatial coordinates were among the most powerful predictors and the modeling approach with physiographic and stand structural variables together was the most improved relative to the average frequency of occurrence. GIS probability maps produced with the application of the physiographic models showed good spatial congruence between high probability values and plots that contained CLUN. NPMR proved to be a reliable probability modeling and mapping tool that could be used as the analytical link between monitoring and quantifying the status and trends of vegetation resources." |

References

- Agresti, A. 1990. *Categorical Data Analysis*. Wiley, New York.
- Agterberg, F. P. 1984. Trend surface analysis. Pages 147-171 in G. L. Gaile, & C. J. Willmott (eds.), *Spatial Statistics and Models*. Reide, Dordrecht, The Netherlands.
- Antoine, M. E. & B. McCune. 2004. Contrasting fundamental and realized ecological niches with epiphytic lichen transplants in an old-growth *Pseudotsuga* forest. *Bryologist* 107:163-173.
- Assaid, C. A. & J. B. Birch. 2000. Automatic bandwidth selection in robust nonparametric regression. *Journal of Statistical Computing and Simulation* 66:259-272.
- Austin, M. P. 1987. Models for the analysis of species response to environmental gradients. *Vegetatio* 69:35-45.
- Austin, M. P. 2002. Spatial prediction of species distribution: an interface between ecological theory and statistical modeling. *Ecological Modeling* 157:101-118.
- Austin, M. P., R. B. Cunningham & P. M. Flemming. 1984. New approaches to direct gradient analysis using environmental scalars and statistical curve-fitting procedures. *Vegetatio* 55:11-27.
- Austin, M. P., A. O. Nicholls & C. R. Margules. 1990. Measurement of the realized qualitative niche: environmental niches of five *Eucalyptus* species. *Ecological Monographs* 60:161-177.
- Austin, M. P., A. O. Nicholls, M. D. Doherty & J. A. Meyers. 1994. Determining species response functions to an environmental gradient by means of a beta-function. *Journal of Vegetation Science* 5:215-228.
- Berryman, S. & B. McCune. 2006. Estimating epiphytic macrolichen biomass from topography, stand structure and lichen community data. *Journal of Vegetation Science* 17:157-170.
- Binder, M.B. & Ellis, C.J. 2008. Conservation of the rare British lichen *Vulpicida pinastri*: climate change, habitat loss and strategies for mitigation. *Lichenologist* 40: 63-79.
- Bio, A. M. F., R. Alkemade, & A. Brarendregt. 1998. Determining alternative models for vegetation response analysis: a non-parametric approach. *Journal of Vegetation Science* 9:5-16.
- Bowman, A. W. and A. Azzalini. 1997. *Applied smoothing techniques for data analysis*. Clarendon Press, Oxford. 193 pp.
- Boyce, M. S., P. R. Vernier, S. E. Nielsen & F. K. A. Schmiegelow. 2002. Evaluating resource selection functions. *Ecological Modelling* 157:281-300.
- Cade, B. S. & B. R. Noon. 2003. A gentle introduction to quantile regression for ecologists. *Frontiers in Ecology and the Environment* 1:412-420.
- Casazza,, G., E. Zappa, M. G. Mariotti, F. Médail & L. Minuto. 2007.2008. Ecological and historical factors affecting distribution pattern and richness of endemic plant species: the case of the Maritime and Ligurian Alps hotspot. *Diversity and Distributions* 14: 47 – 58.
- Clark, R. M. 1975. A calibration curve for radio carbon dating. *Antiquity* 49:251-266.
- Cleveland, W. S. 1979. Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association* 74:829-836.
- Crabtree, D. & C. J. Ellis. 2010. Species interaction and response to wind speed alter the impact of projected temperature change in a montane ecosystem. *Journal of Vegetation Science* 21:744-760.
- Cristofolini, F., P. Giordani, E. Gottardini, & Paolo Modenesi. 2008. The response of epiphytic lichens to air pollution and subsets of ecological predictors: A case study from the Italian Prealps. *Environmental Pollution* 151: 308-317.
- Curtis, J. T. 1959. *The Vegetation of Wisconsin*. University of Wisconsin Press, Madison.
- DeBano, S. J., P. B. Hamm, A. Jensen, S. I. Rondon. and P. J. Landolt. 2010. Spatial and temporal dynamics of potato tuberworm (Lepidoptera: Gelechiidae) in the Columbia Basin of the Pacific Northwest. *Environ. Entomol.* 39:1-14.
- Derr, C. C., B. McCune & L. H. Geiser. 2007. Epiphytic macrolichen communities in *Pinus contorta* peatlands in southeastern Alaska. *Bryologist* 110: 521–532.
- Duren, O. C. & P. S. Muir. 2010. does fuels management accomplish restoration in southwest Oregon, USA, chaparral? Insights from age structure. [Fire Ecology](#) 6(2):76-96. doi: 10.4996/fireecology.0602076
- Ellis, C.J. & Coppins, B.J. 2007. Changing climate and historic-woodland structure interact to control species diversity of the ‘Lobarion’ epiphyte community in Scotland. *Journal of Vegetation Science* 18: 725-734.

- Ellis, C. J., B. J. Coppins & T. P. Dawson. 2007a. Predicted response of the lichen epiphyte *Lecanora populicola* to climate change scenarios in a clean-air region of Northern Britain. *Biological Conservation* 135: 396-404.
- Ellis, C. J., B. J. Coppins, T. P. Dawson, & M. R.D. Seaward. 2007b. Response of British lichens to climate change scenarios: Trends and uncertainties in the projected impact for contrasting biogeographic groups. *Biological Conservation* 140:217-235
- Erjinaes, R. 2000. Can we trust gradients extracted by detrended correspondence analysis? *Journal of Vegetation Science* 11:573-584.
- Eubank, R. L. 1999. *Nonparametric Regression and Spline Smoothing*. 2nd ed. Marcel Dekker, Inc., New York. 338 pp.
- Fan, J. 1993. Local linear regression smoothers and their minimax efficiency. *Annals of Statistics* 21:196-216.
- Fan, J. & I. Gijbels. 1996. *Local Polynomial Modelling and Its Applications*. Chapman & Hall, London. 341 pp.
- Fenton, N. J. & Y. Bergeron. 2008. Does time or habitat make old-growth forests species rich? Bryophyte richness in boreal *Picea mariana* forests. *Biological Conservation* 141: 1389-1399.
- Fielding, A. H. and Bell, J. F. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation* 24:38-49.
- Fleishman, E., R. MacNally, & J. P. Fay. 2003. Validation tests of predictive models of butterfly occurrence based on environmental variables. *Conservation Biology* 17:806-817.
- Flitcroft, R. L. 2008. Regions to streams : spatial and temporal variation in stream occupancy patterns of coho salmon (*Oncorhynchus kisutch*) on the Oregon coast. PhD Dissertation, Oregon State University, 206 pp.
- Fox, J. 2002. *Nonparametric Regression*. Appendix to an R and S-PLUS Companion to Applied Regression. 15 pages.
- Franklin, J. 1995. Predictive vegetation mapping: geographic modelling of biospatial patterns in relation to environmental gradients. *Prog. Phys. Geogr.* 19:474-499.
- Franklin, J. 1998. Predicting the distribution of shrub species in southern California from climate and terrain-derived variables. *Journal of Vegetation Science* 9:733-748.
- Friedman, J.H. 1991. Multivariate adaptive regression splines (with discussion). *Ann. Stat.* 19:1-141.
- Friedman, J. H. and C. B. Roosen. 1995. An introduction to multivariate adaptive regression splines. *Statistical Methods in Medical Research* 4:197-217.
- Gignac, L. D., D. H. Vitt, S. C. Zoltai, & S. E. Bayley. 1991a. Bryophyte response surfaces along climatic, chemical, and physical gradients in peatlands of western Canada. *Nova Hedwigia* 53:27-71.
- Gignac, L. D., D. H. Vitt & S. E. Bayley. 1991b. Bryophyte response surfaces along ecological and climatic gradients. *Vegetatio* 93:29-45.
- Gilligan, L. A. & P. S. Muir. 2011. Stand structures of Oregon white oak woodlands, regeneration, and their relationships to the environment in southwestern Oregon. *Northwest Science* 85:141-158.
- Giordani, P. 2007. Is the diversity of epiphytic lichens a reliable indicator of air pollution? A case study from Italy *Environmental Pollution* 146: 317-323.
- Giordani, P. & G. Incerti. 2008. The influence of climate on the distribution of lichens: a case study in a borderline area (Liguria, NW Italy). *Plant Ecology* 195: 257-272.
- Glavich, D. 2009. Distribution, rarity and habitats of three aquatic lichens on federal land in the U.S. Pacific Northwest. *Bryologist* 112:54-72.
- Gotway, C. A., R. B. Ferguson, G. W. Hergert, & T. A. Peterson. 1996. Comparison of kriging and inverse-distance methods for mapping soil parameters. *Soil Science Society of America Journal* 60:1237-1247.
- Green, P. J. & B. W. Silverman. 1994. *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach*. Chapman and Hall, London. 182 pp.
- Grundel, R. & N. B. Pavlovic. 2007. Response of bird species densities to habitat structure and fire history along a Midwestern open-forest Gradient. *The Condor* 109:734-749
- Guisan, A. & N. E. Zimmermann. 2000. Predictive habitat distribution models in ecology. *Ecological Modelling* 135:147-186.
- Hardle, W. 1990. *Applied Nonparametric Regression*. Cambridge University Press, Cambridge.
- Hardle, W., P. Hall, and J. S. Marron. 1988. How far are automatically chosen regression smoothing parameters from their optimum. *Journal of the American Statistical Association* 83:86-95.

- Hardle, W., P. Hall, and J. S. Marron. 1992. Regression smoothing estimators that are not far from their optimum. *Journal of the American Statistical Association* 87:227-233.
- Harrell, F. E., K. L. Lee, & D. B. Mark. 1996. Multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat. Med.* 15:361-387.
- Hastie, T.J. & Tibshirani, R.J. 1990. *Generalized Additive Models*. Chapman and Hall, London.
- Hastie, T., Tibshirani, R. & Friedman, J. 2001. *The Elements of Statistical Learning*. Springer-Verlag, New York.
- Heegaard, E. 2002a. The outer border and central border for species-environment relationships estimated by non-parametric generalized additive models. *Ecological Modelling* 157:131-139.
- Heegaard, E. 2002b. A model of alpine species distribution in relation to snowmelt time and altitude. *Journal of Vegetation Science* 13:493-504.
- Hirzel, A. H., J. Hausser, D. Chessel, and N. Perrin. 2002. Ecological-niche factor analysis: how to compute habitat suitability maps without absence data. *Ecology* 83:2027-2036.
- Hosten, P. E., H. Whitridge, D. Schuster, and J. Alexander. 2007. Livestock on the Cascade-Siskiyou National Monument: A Summary of Stocking Rates, Utilization, and Management. U.S. Department of the Interior, Bureau of Land Management, Medford District.
- Huisman, J., H. Olff & L. F. M. Fresco. 1993. A hierarchical set of models for species response analysis. *Journal of Vegetation Science* 4:37-46.
- Huntley, B., P. J. Bartlein, & I. C. Prentice. 1989. Climatic control of the distribution and abundance of beech (*Fagus L.*) in Europe and North America. *Journal of Biogeography* 16:551-560.
- Huntley, B., P. M. Berry, W. Cramer, & A. P. McDonald. 1995. Modelling present and potential future ranges of some European higher plants using climate response surfaces. *Journal of Biogeography* 22:967-1001.
- Hurvich, C. M., J. S. Simonoff, and C.-L. Tsai. 1998. Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion. *Journal of the Royal Statistical Society B.* 60, part 2:271-293.
- Huston, M. A. 2002. Critical issues for improving predictions. Pages 7-21 in J. M. Scott, P. J. Heglund, M. L. Morrison, J. B. Haufler, M. G. Raphael, W. A. Wall, & F. B. Samson, eds., *Predicting Species Occurrences: Issues of Accuracy and Scale*. Island Press, Washington.
- Hutchinson, G. E. 1957. Concluding remarks. *Cold Spring Harbor Symposia on Quantitative Biology* 22:415-427.
- Hutchinson, G. E. 1965. The niche: an abstractly inhabited hypervolume. Pages 26-78 in *The Ecological Theatre and the Evolutionary Play*. Yale University Press, New Haven, Conn.
- Irvine, D. R., D. E. Hibbs, and J. P. A. Shatford. 2009. The relative importance of biotic and abiotic controls on young conifer growth after fire in the Klamath-Siskiyou region. *Northwest Science* 83:334-347.
- Jovan, S. 2003. Distribution and habitat models of epiphytic *Physconia* in north-central California. *Bulletin of the California Lichen Society* 10:29-35.
- Jovan, S. & B. McCune. 2005. Air-quality bioindication in the greater Central Valley of California, with epiphytic macrolichen communities. *Ecological Applications* 15:1712-1726.
- Jovan, S. & B. McCune. 2006. Using epiphytic macrolichen communities for biomonitoring ammonia in forests of the greater Sierra Nevada, California. *Water, Air and Soil Pollution* 170:69-93.
- Kaiser, M. S., P. L. Speckman, & J. R. Jones. 1994. Statistical models for limiting nutrient relations in inland waters. *Journal of the American Statistical Association* 89: 410-423.
- Kass, R. E. and Raftery, A. E. 1995. Bayes factors. *Journal of the American Statistical Association* 90:773-795.
- Kohler, G. R. 2007. Predators associated with hemlock woolly adelgid (Hemiptera: Adelgidae) infested western hemlock in the Pacific Northwest. M.S. Thesis, Oregon State University. 135 pp.
- Leathwick, J. R. 1995. Climatic relationships of some New Zealand forest tree species. *Journal of Vegetation Science* 6:237-248.
- Lintz, H. E., B. McCune, A. N. Gray, K. A. McCulloh. 2011. Quantifying ecological thresholds from response surfaces. *Ecological Modelling* 222: 427-436. doi:10.1016/j.ecolmodel.2010.10.017
- Loucks, O. L. 1962. Ordinating forest communities by means of environmental scalars and phytosociological indices. *Ecological Monographs* 32:137-166.
- Martinez-Taberner, A., M. Ruiz-Perez, I. Mestre, & V. Forteza. 1992. Prediction of potential submerged vegetation in a silted coastal marsh, Albufera of Majorca, Balearic Islands. *Journal of Environmental Management* 35:1-12.

- McCune, B. 2006. Nonparametric habitat models with automatic interactions. *Journal of Vegetation Science* 17: 819-830.
- McCune, B. 2007. Improved estimates of incident radiation and heat load using nonparametric regression against topographic variables. *Journal of Vegetation Science* 18: 751-754.
- McCune, B., S. D. Berryman, J. H. Cissel, and A. I. Gitelman. 2003. Use of a smoother to forecast occurrence of epiphytic lichens under alternative forest management plans. *Ecological Applications* 13:1110-1123.
- McCune, B. and J. B. Grace. 2002. *Analysis of Ecological Communities*. MjM Software, Gleneden Beach, Oregon, USA (www.pcord.com) 304 pages.
- McCune, B., S. Jovan and A. Hardman. 2008. Changes in forage lichen biomass after insect outbreaks and fuel reduction treatments in the Blue Mountains, Oregon. *North American Fungi* 3(4): 1-15.
- McCune, B. and Mefford. 1999. *Multivariate analysis on the PC-ORD system*. Version 4. MjM Software, Gleneden Beach, Oregon, U.S.A.
- McCune, B. and M. J. Mefford. 2004. *HyperNiche. Multiplicative Habitat Modeling*. Version 1. MjM Software, Gleneden Beach, Oregon, U.S.A.
- Miller, S. W., D. Wooster, & J. Li. 2007. Resistance and resilience of macroinvertebrates to irrigation water withdrawals. *Freshwater Biology* 52: 2494-2510.
- Minuto, L., F. Grassi & G. Casazza. 2006. Ecogeographic and genetic evaluation of endemic species in the Maritime Alps: the case of *Moehringia lebrunii* and *M. sedoides* (Caryophyllaceae). *Plant Biosystems* 140: 146-155.
- Morrison, M. L. & L. S. Hall. 2002. Standard terminology: toward a common language to advance ecological understanding and application. Pages 43-52 in J. M. Scott, P. J. Heglund, M. L. Morrison, J. B. Haufler, M. G. Raphael, W. A. Wall, & F. B. Samson, eds., *Predicting Species Occurrences: Issues of Accuracy and Scale*. Island Press, Washington.
- Neter, J., M. H. Kutner, C. J. Nachtsheim, & W. Wasserman. 1996. *Applied Linear Statistical Models*. Fourth Edition. McGraw-Hill, Boston. 1408 pp.
- Oksanen, J., E. Läära, P. Huttunen & J. Meriläinen. 1988. Estimation of pH optima and tolerances of diatoms in lake sediments by the methods of weighted averaging, least squares and maximum likelihood, and their use for the prediction of lake acidity. *Journal of Paleolimnology* 1:39-49.
- Oksanen, J. 1997. Why the beta-function cannot be used to estimate skewness of species responses. *Journal of Vegetation Science* 8:147-152.
- Oreskes, N., K. Shrader-Frechette, & K. Belitz. 1994. Verification, validation, and confirmation of numerical models in the earth sciences. *Science* 263:641-646.
- Pawitan, Y. 2001. *In All Likelihood. Statistical Modelling and Inference Using Likelihood*. Clarendon Press, Oxford. 528 pp.
- Peters, R. H. 1991. *A Critique for Ecology*. Cambridge University Press, Cambridge.
- Peterson, E. B. 2000. *Analysis and Prediction of Patterns in Lichen Communities over the Western Oregon Landscape*. Ph.D. Dissertation, Oregon State University, Corvallis.
- Peterson, A. T. 2001. Predicting species' geographic distributions based on ecological niche modeling. *Condor* 103:599-605.
- Peterson, A. T. & C. R. Robins. 2003. Using ecological-niche modeling to predict barred owl invasions with implications for spotted owl conservation. *Conservation Biology* 17:1161-1165.
- Peterson, A. T., D. R. B. Stockwell, & D. A. Kluza. 2002. Distributional prediction based on ecological niche modeling of primary occurrence data. Pages 617-623 in J. M. Scott, P. J. Heglund, & M. L. Morrison, editors. *Predicting species occurrences: issues of scale and accuracy*. Island Press, Washington, D.C.
- Ponadera, K. C. & M. G. Potapova. 2007. Diatoms from the genus *Achnantheidium* in flowing waters of the Appalachian Mountains (North America): Ecology, distribution and taxonomic notes. *Limnologia. Ecology and Management of Inland Waters* 37: 227-241.
- Ponzetti, J., B. McCune & D. A. Pyke. 2007. Biotic soil crusts in relation to topography, cheatgrass and fire in the Columbia Basin, Washington. *Bryologist* 110: 706-722.
- Potapova, M. G. & D. M. Wintel. 2006. Use of nonparametric multiplicative regression for modeling diatom habitat: a case study of three *Geissleria* species from North America. *Advances In Phycological Studies, Festschrift in Honour of Prof. Dobrina Temniskova-Topalova*, pp. 319-332.

- Reusser, D. A. and H. Lee, II. 2008. Predictions for an invaded world: a strategy to predict the distribution of native and non-indigenous species at multiple scales. *ICES Journal of Marine Science* 65: 742-745.
- Rood, S. B., L. A. Goater, K. M. Gill and J. H. Braatne. 2010. Sand and sandbar willow: a feedback loop amplifies environmental sensitivity at the riparian interface. *Oecologia* 165: 31-40.
- Scachetti-Pereira, R. 2002. Desktop GARP. University of Kansas Natural History Museum, Lawrence, Kansas. <http://www.lifemapper.org/desktopgarp/>
- Schroeder, T. A., A. Hamann, T. Wang & N. C. Coops. 2010. Occurrence and dominance of six Pacific Northwest conifer species. *Journal of Vegetation Science* 21:586-596.
- Scott, D. W. 1992. *Multivariate Density Estimation: Theory, Practice, and Visualization*. John Wiley, New York. 317 pp.
- Scott, J. M., P. J. Heglund, M. L. Morrison, J. B. Haufler, M. G. Raphael, W. A. Wall, & F. B. Samson. 2002. *Predicting Species Occurrences: Issues of Accuracy and Scale*. Island Press, Washington. 868 pp.
- Sokal, R. R. and Rohlf, F. J. 1995. *Biometry*. 3rd edition. W. H. Freeman, New York.
- Vander Haegen, W. M., M. A. Schroeder, S. S. Germaine, S. D. West & R. A. Gitzen. 2004. Wildlife on Conservation Reserve Program lands and native shrubsteppe in Washington. Progress Report: 2004 http://wdfw.wa.gov/wlm/research/papers/shrub/conservation_reserve_program.pdf
- Vayssières, M. P., R. E. Plant & B. H. Allen-Diaz. 2000. Classification trees: an alternative non-parametric approach for predicting species distributions. *Journal of Vegetation Science* 11:679-694.
- Wand, M. P. & M. C. Jones. 1995. *Kernel Smoothing*. Chapman & Hall, London.
- Wedderburn, S. D., K. F. Walker, & B. P. Zampatti. 2007. Habitat separation of *Craterocephalus* (Atherinidae) species and populations in off-channel areas of the lower River Murray, Australia. *Ecology of Freshwater Fish* 16: 442-449. <http://dx.doi.org/10.1111/j.1600-0633.2007.00243.x>
- Welsh, H. H., Jr. & G. R. Hodgson. 2010. Spatial relationships in a dendritic network: the perpetofaunal metacommunity of the Mattole River catchment of northwest California. *Ecography* 33:1-18.
- Welsh, H. H., Jr., G. R. Hodgson, J. J. Duda & J. M. Emlen. 2010. Faunal assemblages and multi-scale habitat patterns in headwater tributaries of the South Fork Trinity River – an unregulated river embedded within a multiple-use landscape. *Animal Biodiversity and Conservation* 33: 63-87.
- Wiens, J. A. 1989. *The Ecology of Bird Communities*. Vol. 1, Foundations and Patterns. Cambridge University Press, Cambridge.
- Whittaker, R. H. 1954. Plant populations and the basis of plant identification. In: *Augewandte Pflanzensoziologie, Veröffentlichungen des Karntner Landesinstituts für angewandte Pflanzensoziologie in Klagenfurt*, Festschrift Aichinger, Vol. 1.
- Whittaker, R. H. 1956. Vegetation of the Great Smoky Mountains. *Ecological Monographs* 26:1-80.
- Yee, T. W. & N. D. Mitchell. 1991. Generalized additive models in plant ecology. *Journal of Vegetation Science* 2:587-602.
- Yost, A. 2006. Probabilistic modeling of understory vegetation species in a Northeastern Oregon industrial forest. PhD Diss., Oregon State University. 191 pp.
- Yost, A. C. 2008. Probabilistic modeling and mapping of plant indicator species in a Northeast Oregon industrial forest, USA. *Ecological Indicators* 8: 46-56.
- Zaniewski, A. E., A. Lehmann, & J. M. Overton. 2002. Predicting species spatial distributions using presence-only data: a case study of native New Zealand ferns. *Ecological Modelling* 157:261-280.