D. Bayesian probability models. One logical extension of pattern recognition models is based upon the use of Bayesian probabilities. This technique, which eliminates the complexity of having to establish different answer patterns, has been described by Russell et al. (1980) as PATREC models and involves answering the following questions:

(1) What are the highest long-term population densities of this species that can be found in the geographic unit? Either recent or historic densities are appropriate. Record this number as the high density standard.

(2) What are the lowest (but not zero) long-term population densities of this species that can be found in this geographic unit? Record this number as the low density standard.

(3) What is the mid-point of the range between the lowest and highest population densities?

$$\frac{\text{Low} + \text{High}}{2} = \text{Midpoint} \tag{19}$$

(4) What percent of the total habitat within the geographic unit supports population densities higher than the midpoint of the range of densities calculated earlier? Record this number as the prior probability for High.

(5) What is the difference between 1.0 and the prior probability for High? Record this number as the prior probability for Low.

(6) Are any of the variables identified in the variable identification phase of modeling (103 ESM 3.2) related to each other (e.g., are any of the variables directly a consequence of or an extension of any other variable)?

(7) If the answer is yes for one or more variables in the preceding step, which of the two do you think has a more direct influence on population density, or is more readily measured? Remove the other related variables from the list. If the answer is no, continue with question (8).

(8) Which, if any, of the variables remaining on the list have no apparent diagnostic value (e.g., is present in about the same amount in locations where the population density is low as well as where it is high)? Remove any such variables from the list.

(9) Beginning with any variable on the list, in what unit is the variable measurable? Choices include English units, metric units, percent, degrees (e.g., compass, slope, temperature), and presence or absence.

(10) What is the range of possible measurements for the variable (e.g., the range of possible measurements for dbh is 0-100 cm). Within this range, are there subranges where the variable has a different effect on population density (e.g., a dbh of greater than 50 cm is optimum for red-tailed hawk nesting)? Record each range and label as a, b, etc. (e.g., a = dbh of 0-50 cm; b = dbh of greater than 50 cm).

(11) What is the relative frequency with which subrange increment "a" is associated with a high population density (e.g., what percent of high density areas have trees of average dbh between 0 and 50 cm)?

(12) What is the relative frequency with which subrange increment "a" is associated with a low population density (e.g., what percent of low density areas have trees of average dbh between 0 and 50 cm)?

(13) Repeat questions 11 and 12 for each subrange increment.

(14) For all subranges, the relative frequencies in the "High" column must sum to exactly 1.0. If not, adjust the relative frequencies until they equal 1.0.

(15) For all subranges, the relative frequencies in the "Low" column also must sum to exactly 1.0. If not, adjust the relative frequencies until they equal 1.0.

(16) Select the next variable and repeat questions 9 through 15. Continue this procedure until all of the variables (and subranges of their values) have been assigned relative frequencies for both high and low population areas.

An example PATREC model is shown in Table 3-7.

Table 3-7. Example PATREC model for the red-tailed hawk.

Variable	Measurement	Subrange	Probability Given High Population	Probability Given Low Population
1. % herbaceous canopy	≥70%	a	0.6	0.2
	<70%	b	0.4	0.8
2. Proportion of herbaceous canopy between 8 and 46 cm	≥50%	a	0.7	0.1
	<50%	b	0.3	0.9
		Prior Probabilities:	0.6	0.4

The above PATREC model is based upon the following:

(1) Prior probabilities are assumed to be: High = 0.60; Low = 0.40.

(2) The variables (and their subranges) are those depicted in Table 3-7.

(3) The high density standard is 20 hawks/10 km^2 and the low density standard is 1 hawk/10 km^2.

Application of the PATREC model is illustrated by the following example. Suppose that the two variables are measured in a hypothetical study area. The average values are 50% herbaceous canopy cover (subrange b of Variable 1) and 50% of herbaceous canopy between 8 and 46 cm (subrange a of Variable 2). The subranges for each variable identify probability estimates in both the high population and low population columns of Table 3-7. For both the high and low probability columns, the identified probabilities are multiplied together with the prior probability as follows:

348-753 O - 81 - 5

Product of high = (0.4)(0.7)(0.6) = 0.168 (20)

Product of low = (0.8)(0.1)(0.4) = 0.032 (21)

The probability of a high population potential is calculated by dividing the product of high (equation 20) by the sum of the products (equation 20 and 21) as follows:

$$\text{Probability of high} = \frac{0.168}{0.168 + 0.032} = 0.84 \quad (22)$$

The probability of a low population potential is computed by dividing the product of low (equation 21) by the same sum as follows:

$$\text{Probability of low} = \frac{.032}{0.168 + 0.032} = 0.16 \quad (23)$$

(note: probability of low = 1.0 - probability of high)

The predicted population potentially supported by the study area is computed by the following equation:

Potential population = (High Density Standard) (Prob. of high) + (Low Density Standard) (Prob. of low) (24)

= 20(.84) + 1(.16)
= 16.8 + .16
= 16.96

The result from the PATREC calculations indicates a potential density of 16.96/10 km^2. This output from a PATREC model, although expressed in units of population density, is not a prediction of actual populations. It is an expression of the habitat conditions as defined by equation (2). This figure is converted to an HSI by dividing by the high density standard (standard of comparison) as described in 103 ESM 2. The HSI would be 16.96/20 or 0.85.

The proper use of PATREC models requires a knowledge of methods for establishing conditional probabilities (Questions 10, 11, and 12 above). The most desired method for establishing probabilities is by actual field observations. Conditional probabilities established subjectively in the absence of field work may be incorrect because of the complex nature of conditional probabilities. For further details on the use of PATREC, refer to Kling (1980) and Russell et al. (1980).

E. Multivariate statistical models. Values such as standing crop (or biomass) or expert ratings are often predictable using statistical models which are based upon sets of easily measured habitat variables. Usually, a large number of habitat variables are defined in Phase II of model construction (103 ESM 3.2). Statistical methods are used to mathematically determine which variables have the greatest influence on habitat potential as defined, for example, by standing crop. A condensed set of variables can then be used to predict potential standing crop or biomass values. Model outputs which predict standing crop or biomass can be converted to HSI by using equation (25).

$$\text{HSI} = \frac{\text{Predicted Standing Crop (or biomass)}}{\text{Maximum Observed Standing Crop value}} \qquad (25)$$

There are at least two statistical techniques which can be used to construct HSI models for an evaluation species: (1) regression; and (2) discriminant analysis.

(1) Regression models. Standing crop or biomass of a species is often predictable based upon a set of measurable habitat variables. For example, height and density of forest canopy, tree size, percent coniferous cover, and number of shrubs per acre were found by researchers to greatly influence the density of male Acadian flycatchers during the nesting season. These and other habitat variables were used to produce a multiple regression equation (equation (26)) which accurately predicted population densities of seven avian species (Robbins 1978).

$$N = 20.816 + (-2.349\ V_1) + (-0.341\ V_2) + (0.985\ V_3) + (1.988\ V_4) + (-0.000371\ V_5) + (-0.286\ V_6) + (-0.746\ V_7) \qquad (26)$$

where N = male breeding density of Acadian flycatchers;

V_1 = degrees of latitude north or south of 38° after correcting for elevation; departure always given as a positive number;

V_2 = percent of canopy cover expressed as a whole number;

V_3 = mean canopy height in feet;

V_4 = number of trees per acre greater than 21 inches (53.3 cm) dbh;

V_5 = estimated number of shrub stems per acre at breast height;

V_6 = percent of coniferous trees (3 inches dbh or greater) expressed as a whole number; and

V_7 = percent of standing dead trees (3 inches dbh or greater).

Determination of HSI by use of regression models involves two steps: (1) developing a regression equation which predicts the value of interest (e.g., standing crop or biomass for the study area); and (2) converting the standing crop estimate to HSI by use of equation (25) above. Developing a regression equation is dependent upon the availability of adequate data which includes estimates of standing crop or biomass for a number of areas from which corresponding habitat measures also were collected. Two good references covering statistical techniques for developing regression models are Draper and Smith (1966) and Freese (1974).

(2) Discriminant analysis models. Measures of species density or expert ratings could be predicted with a point estimate by use of a regression equation as described above or predicted with a "group" estimate by use of discriminant analysis. For an evaluation species, discriminant functions can be used to predict the population group to which a particular study area belongs (e.g., Table 3-8).

Table 3-8. Example set of population groups for use in a model developed by discriminant analysis.

Population Category	Group Number
Very high (76-80 animals/ha)	4
High (50 to 75 animals/ha)	3
Medium (25 to 49 animals/ha)	2
Low (1 to 24 animals/ha)	1
None (0 animals/ha)	0

The use of discriminant analysis is based upon the assumption that certain habitat characteristics are unique to each population group and thus can be used to predict whether or not the population in a particular study area is likely to occur within that group.

There are two steps for computing HSI using discriminant analysis: (1) develop a set of discriminant functions for predicting group membership; and (2) convert the discriminant model output to HSI. Nie et al. (1975) provided a good, general description of discriminant analysis. For a more technical discussion, see Morrison (1976). Conversion of a discriminant model output to an HSI is accomplished by dividing the model output by the group number representing the "best" group (Equation (27)).

$$HSI = \frac{\text{Number Representing Group the Study Area is Most Likely to Fall In}}{\text{Number Representing the Optimum Group}}$$

$$= 1/4 \qquad (27)$$
$$= 0.25$$

An example showing the development and calculations involved in a discriminant model is provided in Appendix F.

Both regression and discriminant analysis models are based upon empirical observation or statistical inference concerning habitat variables rather than from theoretical relationships. Statistical models often lack causal structure and may be difficult to modify in an operational situation. Habitat changes related to variables that were not statistically important or ranges of variables in model construction beyond those considered may lead to inadequate predictions.

3.4 Phase IV: Document the model. The fourth phase in model development is to document the HSI model, its structure and assumptions, and the complete set of steps needed to implement the model. Documentation is important because: (1) the user must understand the model, its objectives, its basic biological assumptions, and the basis for its construction; (2) documentation of a model aids in the understanding of how habitat is used by the evaluation species; (3) the user should know what to expect from the model; and (4) documentation provides the basis for understanding how the model may be adapted to other applications.

Two levels of documentation take place concurrently throughout HSI model construction. One level represents the gathering of habitat use information about a particular evaluation species, information which ultimately will be used to construct the HSI model. The second level of documentation describes how the species-habitat information is used to construct the HSI model and the steps for using the model. This also describes the conditions under which the model is applicable.

A. Document habitat use information. Species-habitat relationships and life history information are the basic data sources used for developing the HSI model. Documentation should provide a characterization of the general and specific habitat requirements of the evaluation species. The following habitat requirements may be described in model documentation when building the HSI model:

(1) General habitat requirements

(a) Distribution. Brief information on the general distribution of the species or subspecies, and the specific distribution of the species within a given region (e.g., restricted to northern half of region or elevation above 1,500 m).

(b) Cover types in which the species may be found and the resources (e.g., food and cover) provided by each cover type.

(2) Specific habitat requirements. List quantitative and qualitative data on the habitat requirements of the species. Emphasize any seasonal or life stage differences. Example categories of habitat requirements include the following:

(a) Food requirements. Describe the major food items consumed by the species, the foods of young when different from that of adults, and seasonal food preferences.

(b) Water requirements. Describe the dependence of the species on free water for drinking, seasonal differences in water requirements, or water quality characteristics which may be significant to the species.

(c) Reproductive requirements. Describe the specialized habitat needs associated with reproductive activities (e.g., nest trees with dbh greater than 25 cm or gravel bottom for spawning).

(d) Interspersion or home range requirements. List the optimal mix of life requisites, cover types, or life stages, and the home range sizes. Include seasonal differences if appropriate.

B. Document model construction and use. Model documentation is used to substantiate decisions made in the construction process and to identify necessary assumptions and why they were made. It also may be important to document why certain types of information or existing models were not used.

The proper use of a model is largely dependent on adequate instructions for model application and a knowledge of the assumptions and limitations involved in model construction. There are several areas where documentation is advised to ensure proper model construction and use:

(1) Information pertaining to the applicability of the model should be clearly defined. This includes the geographic area of applicability, season of applicability, and the verification level of the model (see 103 ESM 3.5).

(2) The selected variables must be justified, with an explanation of why each is important. Describe how each variable is related to the model components. This task may be accomplished by a graphical display (tree-diagram) and a verbal description of the model (see 103 ESM 3.2).

(3) Assumptions made in defining model relationships should be explained. The method for determining suitability for each variable should be explained when suitability indices are established for model variables. Explanations should be provided to substantiate the aggregation techniques used in the mathematical formulation of the model (see 103 ESM 3.3). For many existing models, instructions for converting the model output to an HSI are necessary. These instructions, along with assumptions and limitations of the conversion process, should be clearly explained.

(4) There should be adequate and complete documentation of steps needed to apply the model, including when, where, and how each variable is measured in the field. The documentation should include detailed instructions for converting study area data collected for each variable into model inputs and instructions for mathematical calculations involved in determining an HSI.

(5) Instructions for interpretation of model outputs should be provided, if appropriate.

3.5 Phase V: Verify the model. The purpose of model verification is to ensure that the model produces an output at the acceptance level established in the first phase, setting model objectives (103 ESM 3.1). Verification serves as a quality check on the model and can be used to further refine or advance a model to a higher acceptance level.

A. Four steps of model verification. Verification includes four progressive steps (Figure 3-17). The degree to which model verification proceeds through these steps depends upon the specified acceptance level.

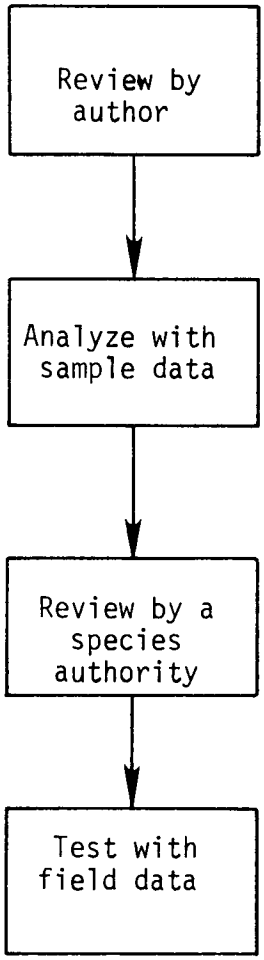

Figure 3-17. Four steps of model verification.

3. Construction of HSI Models

(1) Step 1: Review by author. The author should check model components and the relationships identified in the second modeling phase (Variable Identification, 103 ESM 3.2). The author should ensure that the appropriate variables were used and the relationships between variables were accurately portrayed. The mathematics and graphs used should be checked to make certain that they are consistent with the perceived habitat relationships in the model. The symbols, operations, and signs used in each equation should be verified. It also is important to verify that all model assumptions and limitations were stated correctly. As a result of this review, the model assumptions and variable relationships may need to be reformulated.

(2) Step 2: Analyze with sample data. The model is applied to sample data sets (either real or hypothetical) which mimic various habitat conditions (e.g., high, medium, and low suitability). Inspection of the HSI outputs reveal how well the model reflects the habitat condition for each data set. If necessary the model should be calibrated to give what appears to be a reasonable prediction for each set of conditions.

(3) Step 3: Review by a species authority. The model, its documentation, and the results of the sample data analysis are reviewed by an authority on the species. The objective of this review is to increase the reliability of the model. Any weak or questionable parts of the model should be identified so that improvements can be made. Also, supporting information such as a description of how the model is to be used and the required level of resolution may be helpful in this review.

(4) Step 4: Test with field data. Testing the model against actual field data should occur only after the model has gone through the first three steps of verification. Field data must be available that provide both measurements of habitat variables and measures which represent habitat suitability (e.g., standing crop estimates or habitat ratings by a species authority). Statistical expertise is required to design and perform the appropriate test(s).

Tests utilizing field data may lead to a refinement or updating of the model because more habitat data may be obtained in the tests than were available for initial model construction. Updating the model could mean that the model is revised to meet a new acceptance level.

B. Application of the verification steps. The four verification steps described above are used to ensure that the model meets the desired level of acceptance. However, all four steps need not be applied in every situation. The specified acceptance level defines which of the verification steps are to be performed in a model test. For example, Table 3-9 displays which verification steps should be performed for three potential acceptance levels established for an HSI model.

Table 3-9. Verification steps required for various acceptance levels.

	Acceptance level	Verification steps to be applied
1.	The model appears to be reasonable to an evaluation team (the author).	1, 2
2.	The model appears to be reasonable to a species authority	1, 2, 3
3.	The model is applied at several sites and the HSI tends to be correlated with standing crop estimates at each site.	1, 2, 3, 4

C. Summary. The results of the verification process should be well documented to ensure that the utility of the model does not fall short of preconceived expectations and that future model enhancements benefit from past testing. A model can be applied once it has been shown to operate at the desired acceptance level. Future applications may require that the model meet a higher acceptance level. In this case, the verification process is repeated at the higher level but should benefit from prior documentation.

If the acceptance level cannot be met, the problem(s) with the model must be diagnosed. If the problem(s) are not readily solvable, the model's acceptance level can be lowered, and the model applied at the

lower acceptance level. If problems with the model are not solvable and the acceptance level cannot be lowered, the model should not be applied.

The modeling approaches discussed in 103 ESM 3.3 are relatively simplistic and seem appropriate for assessment level habitat models based on currently available data. Because of the general lack of data for most fish and wildlife species, modelers must often rely on intuition or professional interpretation of basic habitat data. In such circumstances, the reliability of a model may be more limited by the lack of data than our ability to describe complex relationships. Realizing these basic limitations, testing should not continue past the point where further improvements can be made in the model.