ERDC/EL TR-01-31

US Army Corps
of Engineers®
Engineer Research and
Development Center

Wetlands Research Program

Hydrogeomorphic Approach to Assessing Wetland Functions: Guidelines for Developing Regional Guidebooks

Chapter 7
Verifying, Field Testing, and Validating Assessment Models

James S. Wakeley and R. Daniel Smith

September 2001

Environmental Laboratory

Wetlands Research Program

ERDC/EL TR-01-31
September 2001

Hydrogeomorphic Approach to Assessing Wetland Functions: Guidelines for Developing Regional Guidebooks

Chapter 7
Verifying, Field Testing, and Validating Assessment Models

by James S. Wakeley, R. Daniel Smith

Environmental Laboratory
U.S. Army Engineer Research and Development Center
3909 Halls Ferry Road
Vicksburg, MS 39180-6199

Final report

Prepared for U.S. Army Corps of Engineers
Washington, DC 20314-1000

Under CRWRP Work Unit 32985

Contents

List of Figures

Preface

This chapter in the Guidelines for Developing Regional Guidebooks was authorized by Headquarters, U.S. Army Corps of Engineers (HQUSACE), as part of the Characterization and Restoration of Wetlands Research Program (CRWRP), Work Unit 32985, "Technical Development of HGM." Mr. Dave Mathis was the CRWRP Coordinator at the Directorate of Research and Development, HQUSACE; Ms. Colleen Charles, HQUSACE, served as the CRWRP Technical Monitor's Representative; and Dr. Russell F. Theriot, Environmental Laboratory (EL), U.S. Army Engineer Research and Development Center (ERDC), was the CRWRP Program Manager.

This report was prepared by Dr. James S. Wakeley and Mr. R. Daniel Smith, Wetlands Branch, Ecological Research Division, EL. The study was conducted under the general supervision of Dr. Morris Mauney, Chief, Wetlands Branch; Dr. Conrad Kirby, former Chief, Environmental Resources Division; and Dr. John W. Keeley, former Director, EL. Dr. Edwin A. Theriot is Acting Director, EL.

Some of the ideas presented in Chapter 7 were based on discussions at a 1989 workshop on Habitat Model Validation in Estes Park, CO, organized by the U.S. Fish and Wildlife Service (FWS). The authors wish to thank Messrs. Adrian "Bubba" Farmer and Rick Schroeder of FWS for many years of interesting and helpful interaction. Mr. Lee Baxter, U.S. Bureau of Reclamation, shared some examples of sensitivity analyses done with spreadsheets, and Mr. Jon Campbell, Wetlands Branch, wrote the spreadsheet presented in Figure 7-2. Dr. Mary Davis, Wetlands Branch, provided many useful comments on the manuscript.

At the time of publication of this report, Dr. James R. Houston was Director of ERDC, and COL John W. Morris III, EN, was Commander and Executive Director.

This report should be cited as follows:

Wakeley, J. S., and Smith, R. D. (2001). "Hydrogeomorphic approach to assessing wetland functions: Guidelines for developing regional guidebooks; Chapter 7, Verifying, field testing, and validating assessment models," ERDC/EL TR-01-31, U.S. Army Engineer Research and Development Center, Vicksburg, MS.

7 Verifying, Field Testing, and Validating Assessment Models

Introduction

To many people, model testing is something that is done only after model construction. Perhaps that impression has been reinforced by considering the topic in one of the final chapters of this guide to the development of Regional Guidebooks for the Hydrogeomorphic (HGM) Approach to wetland functional assessment. On the contrary, model testing should be an ongoing and iterative part of the model construction process (Caswell 1976). Development of an HGM assessment model proceeds from forging of the initial conceptual model, through calibration of the model against reference data, to validation or testing for model accuracy. Throughout this process, continual testing and "tweaking" are needed to ensure that the model performs as intended by its developers and will meet the goals of the Assessment Team (A-Team) for accuracy, consistency of output, and ease of application.

Some aspects of model testing are simpler than others. As discussed later in this chapter, a full validation of model accuracy may involve years of intensive research and data gathering, far beyond the interests, capabilities, or responsibilities of most A-Teams. However, a simple test of model logic and sensitivity can be accomplished in less than an hour, and the results can be used immediately to guide further development of the model. For a relatively simple model of one function involving two or three variables and a straightforward aggregation equation (e.g., an arithmetic mean), model logic may be so obvious that formal logic testing may be unnecessary. However, the performance of a complex model, involving many variables and a complicated equation, may no longer be intuitive. Continual testing of model logic and sensitivity may be needed throughout construction.

Most applications of the HGM Approach involve a number of separate models, one for each of the identified functions of the regional wetland subclass. To minimize confusion, the word model in this chapter refers to the model developed for a single function (e.g., Dynamic Subsurface Water Storage) and not to the set of models developed for a regional wetland subclass.

This chapter considers three aspects of model testing: verification, field testing, and validation. For the sake of discussion, it is assumed that the A-Team has completed the development of one or more conceptual assessment models (Chapter 4) and that these models have been calibrated (Chapter 6) using data gathered in reference wetlands (Chapter 3) representing the range of conditions present in the regional wetland subclass. The models have already been subjected to individual peer review and have been critiqued and modified in a workshop of regional wetland experts. Even if some model testing was done earlier during development, final testing is needed before publication of the models as operational drafts. At this point, all models should be subjected to final verification of model logic and sensitivity and field tested for ease of use. In some cases, validation of model accuracy may also be possible and would add considerably to model reliability and user confidence. However, it is anticipated that model validation will be done mainly by third parties after the operational draft models become widely available.

What is verification?

As used in this guidebook, verification is a check of model logic and sensitivity. The goal of verification is to answer the following kinds of questions. In general, does the model perform as envisioned by its developers? Is it sensitive to the kinds and magnitudes of impacts expected for wetlands in the regional subclass? What are the key variables in the model, and do they correspond to the important attributes and processes that are thought to influence the function? Are all variables in the model actually needed or could the model be simplified without much loss of sensitivity? Is the aggregation equation appropriate? Are different variables given appropriate weight in the outcome? Note that model *accuracy* is not an issue here (see section, “What is validation?”). A model that is adequately verified may still be invalid (i.e., give incorrect results).

What is field testing?

Field testing ensures that typical users can apply the model efficiently and with consistent results. One goal of the HGM Approach is the capability to assess wetland functions rapidly, within the time and other constraints imposed by regulatory programs. Therefore, a field test should determine how long it takes to apply the model in typical field situations, identify incomplete or ambiguous instructions, and ensure that the level of training and expertise required to use the model are appropriate. In addition, field testing should verify that the model can be used consistently year-round (if that is what the authors intended) and that different investigators applying the model in the same area get the same results.

What is validation?

Validation is the testing of model accuracy or reliability by comparing model output against an independent measure of the function. The output of an HGM assessment model is a Functional Capacity Index (FCI), which is an index of the ability of the wetland to perform a particular function. For example, one way to validate an assessment model for Particulate Retention in wetlands is by comparing FCI values predicted by the model at a series of wetland sites against a direct measure of sediment accretion in each wetland over a period of time using feldspar clay pads or sediment disks (e.g., Kleiss 1996). How closely FCI values and direct measures of wetland function must coincide for a model to be considered "valid" is up to model developers and users, and may vary with the intended application (Rykiel 1996). A model is an abstraction or approximation of reality and thus can never fully describe the real system. Nonetheless, models are useful because they help us to understand the system and to predict the effects of environmental change (Hall and Day 1977).

The process of model validation is similar to hypothesis testing in statistics. One devises a test and, based on the results, either rejects or fails to reject the model (Caswell 1976; Overton 1977; Marcot, Raphael, and Berry 1983). The model can never be "proven" based on one or more tests; however, confidence in a model increases each time it survives another test.

Although the statistical analogy implies the risk of model rejection, in fact validation should be viewed as part of a continuing process of model modification and improvement (Overton 1977). Model testing is meaningless unless the results are used to improve model performance. A conclusion that "the model doesn't work" is not constructive and will never lead to progress in the science and art of wetland evaluation. Therefore, a proper model validation study should result in the development and testing of an alternative model (O'Neil et al. 1988).

Verifying the Model

HGM assessment models are based initially on available literature and on the experience and judgment of A-Team members. Later, they are calibrated using field data from reference wetlands. In addition, the models are subjected to individual peer review and collective review at a regional workshop that includes wetland experts not involved with the development of the model. Some authors consider peer review to be part of the model verification process (e.g., U.S. Fish and Wildlife Service 1981). For convenience, however, guidelines for peer review and workshop development were presented earlier in this guidebook (see Chapter 1). This section on model verification addresses the following question: Does the model produce logical results?

To verify the logic of an HGM assessment model, one simply applies the model to real or hypothetical data and evaluates the results in light of one's experience and understanding of the regional wetland subclass. Verification is a fairly subjective procedure that is meant to determine whether model output makes sense, and should not be confused with model validation or testing for accuracy (Schroeder and Haire 1993). Model verification can be done on either the conceptual or the calibrated model. There are two basic approaches to testing model logic: (a) performing a sensitivity analysis and (b) applying the model to sample data sets.

Performing a sensitivity analysis

A sensitivity analysis is an appraisal of model performance under incremental change in the input variables (Waide and Webster 1976; Overton 1977). Sensitivity analysis helps to verify that the model will behave as intended under both moderate and extreme levels of each variable (Schroeder and Haire 1993). An important goal of sensitivity analysis is to identify key variables in the model (i.e., those having the most influence on FCI values) and, conversely, those variables that have little influence on model outcome. Variables that do not affect FCI values appreciably should be considered for elimination from the model as a way of reducing sampling effort and enhancing the role of the remaining variables in the model. Alternatively, the A-Team may wish to develop more accurate sampling methods for key variables while relying on more qualitative field methods for the less influential variables.

HGM assessment models are structured as a series of steps leading from field measurements of environmental variables to calculation of the FCI, as follows:

Measures of Model Variables	→	Subindices of Model Variables	→	FCI

Measures of each variable are first converted into subindices (scaled from 0 to 1) based on quantitative relationships defined in the model. Subindices, in turn, are aggregated to determine FCI using a simple, weighted equation that describes how model variables interact to influence the level of function (see Chapter 4).

Sensitivity analyses of HGM assessment models are usually done by inputting different levels of the subindices and examining the effects on FCI. An analysis of this type is useful in verifying that the aggregation equation is working as intended, subindices for each variable are weighted properly, and FCI values are in the proper range (0 to 1). It also is used to determine which variables have the most (or least) influence on model results. However, this kind of analysis will not check that conversions of measures to subindices are appropriate, nor that the model responds as intended to realistic levels of the environmental measurements. Therefore, additional checks are needed to fully verify model logic (see the section "Applying the model to sample data sets").

The easiest way to perform a sensitivity analysis of an assessment model is to enter the aggregation equation into a spreadsheet and incrementally vary the inputs to the model one variable at a time. Effects on FCI can be examined directly from the spreadsheet, or simple statistics (e.g., means, ranges) can be used to quantify the influence of each variable on FCI predictions. More advanced applications use the software's graphing capabilities to plot changes in FCI under different combinations of subindex values.

Figure 7-1 presents a simple sensitivity analysis of a hypothetical three-variable model for the carbon export function of a riverine wetland. The variables are flood frequency V_{FREQ} and abundances of leaf litter V_{LITTER} and coarse woody debris V_{CWD}. The spreadsheet calculates FCI values for all possible combinations of the three variables for subindices equal to 0.0, 0.1, 0.5, and 1.0. Some characteristics of the model are immediately obvious. First, whenever the subindex for V_{FREQ} equals 0, the model always returns an FCI of 0. However, when either V_{LITTER} equals 0 or V_{CWD} equals 0 (but not both), FCI values may range from 0 to 0.71. Therefore, V_{FREQ} has a controlling influence over model output. This form of model may be appropriate if the wetland function simply cannot occur without some important environmental feature or process (e.g., carbon export cannot occur when flood frequency is zero).

Other characteristics of the model shown in the spreadsheet (Figure 7-1) include the fact that FCI = 0.5 when all subindices are set to 0.5, and that FCI = 1.0 only when all the subindices equal 1.0. The A-Team must decide whether the model behaves as the team intended. Use of the spreadsheet easily permits other aggregation equations to be tested until the intended model behavior is achieved. For example, the A-Team may believe that middle-of-the-road values (e.g., 0.5) for all three variables should depress FCI below 0.5. One option to achieve this result is to remove the exponent from the aggregation equation, resulting in FCI = 0.25.

For a complicated model, it may be difficult to interpret model behavior from tabular spreadsheet output alone. Summary statistics and plots of model output are needed. Figure 7-2 presents a sensitivity analysis for a four-variable model that was performed using a set of flexible spreadsheet programs developed especially for this purpose. The programs accept any user-defined model containing up to 15 variables. Spreadsheet files and documentation are available for downloading through the Environmental Laboratory's (EL) HGM Web site at http://www.wes.army.mil/el/wetlands/hgmhp.html. The files are available in Quattro®Pro (*.wb2) and Excel® (*.xls) formats.

The example shown in Figure 7-2 is based on the model for Temporary Storage of Surface Water for low-gradient riverine wetlands in western Kentucky (Ainslie et al. 1999). The aggregation equation is

$$FCI = [(V_{FREQ} \times V_{WIDTH})^{1/2} \times (V_{ROUGH} + V_{SLOPE})/2]^{1/2}$$

The program varies one variable at a time, with all other variables in the model fixed at subindex values of 1, 0.5, 0.1, or 0. The incremented variable is changed from 0 to 1 in increments of 0.1. For example, the first line in the table generated under Step 4 (Figure 7-2) shows that varying V_{FREQ} from 0 to 1, with all other variables fixed at subindices of 1, results in FCI values that range from

V_{FREQ}	V_{LITTER}	V_{CWD}	FCI	V_{FREQ}	V_{LITTER}	V_{CWD}	FCI
0.00	0.00	0.00	0.00	0.50	0.00	0.00	0.00
0.00	0.00	0.10	0.00	0.50	0.00	0.10	0.16
0.00	0.00	0.50	0.00	0.50	0.00	0.50	0.35
0.00	0.00	1.00	0.00	0.50	0.00	1.00	0.50
0.00	0.10	0.00	0.00	0.50	0.10	0.00	0.16
0.00	0.10	0.10	0.00	0.50	0.10	0.10	0.22
0.00	0.10	0.50	0.00	0.50	0.10	0.50	0.39
0.00	0.10	1.00	0.00	0.50	0.10	1.00	0.52
0.00	0.50	0.00	0.00	0.50	0.50	0.00	0.35
0.00	0.50	0.10	0.00	0.50	0.50	0.10	0.39
0.00	0.50	0.50	0.00	0.50	0.50	0.50	0.50
0.00	0.50	1.00	0.00	0.50	0.50	1.00	0.61
0.00	1.00	0.00	0.00	0.50	1.00	0.00	0.50
0.00	1.00	0.10	0.00	0.50	1.00	0.10	0.52
0.00	1.00	0.50	0.00	0.50	1.00	0.50	0.61
0.00	1.00	1.00	0.00	0.50	1.00	1.00	0.71
0.10	0.00	0.00	0.00	1.00	0.00	0.00	0.00
0.10	0.00	0.10	0.07	1.00	0.00	0.10	0.22
0.10	0.00	0.50	0.16	1.00	0.00	0.50	0.50
0.10	0.00	1.00	0.22	1.00	0.00	1.00	0.71
0.10	0.10	0.00	0.07	1.00	0.10	0.00	0.22
0.10	0.10	0.10	0.10	1.00	0.10	0.10	0.32
0.10	0.10	0.50	0.17	1.00	0.10	0.50	0.55
0.10	0.10	1.00	0.23	1.00	0.10	1.00	0.74
0.10	0.50	0.00	0.16	1.00	0.50	0.00	0.50
0.10	0.50	0.10	0.17	1.00	0.50	0.10	0.55
0.10	0.50	0.50	0.22	1.00	0.50	0.50	0.71
0.10	0.50	1.00	0.27	1.00	0.50	1.00	0.87
0.10	1.00	0.00	0.22	1.00	1.00	0.00	0.71
0.10	1.00	0.10	0.23	1.00	1.00	0.10	0.74
0.10	1.00	0.50	0.27	1.00	1.00	0.50	0.87
0.10	1.00	1.00	0.32	1.00	1.00	1.00	1.00

Figure 7-1. Example sensitivity analysis for the model FCI = $[V_{FREQ} \times (V_{LITTER} + V_{CWD})/2]^{1/2}$ done with a spreadsheet

0 to 1. However, varying V_{FREQ} from 0 to 1 with all other variables fixed at 0.5 produces FCIs that range from 0 to 0.59.

The sensitivity analysis in Figure 7-2 shows that the first two variables, V_{FREQ} and V_{WIDTH}, have greater influence over model outcome than either of the other two variables, V_{ROUGH} and V_{SLOPE}. Varying either V_{FREQ} or V_{WIDTH} results in greater change in FCI than does varying either V_{ROUGH} or V_{SLOPE}. Furthermore, the model returns an FCI of 0 when either V_{FREQ} or V_{WIDTH} is 0, but FCI can be as high as 0.71 when either V_{ROUGH} or V_{SLOPE} (but not both) is 0. The effect of incrementing one variable in a complex model is more easily visualized with the

Follow Steps 1-4 to produce a table of FCI ranges.

STEP 1: In spreadsheet cell C4, enter the number of variables in the model you wish to examine.

Num of Variables = 4

STEP 2: Enter these variables by name in Column C, starting in Cell C8.
You can enter between 1 and 15 variables.

V1 --> *Vfreq*
V2 --> *Vwidth*
V3 --> *Vrough*
V4 --> *Vslope*
V5 -->
V6 -->
V7 -->
V8 -->
V9 -->
V10 -->
V11 -->
V12 -->
V13 -->
V14 -->
V15 -->

STEP 3: Enter the equation of the function in cell C24, referencing the cells where the appropriate variables are located from Step 1 above (e.g., @POWER((@POWER((C8*C9),1/2))*((C10+C11)/2),1/2)).

Model:

STEP 4: Press the "Run Macro" button to the right.

Run Macro

Variable Being Incremented from 0 - 1	Subindex Values for Nonincremented Variables											
	1			0.5			0.1			0		
	Range	Low	High	Range	Low	High	Range	Low	High	Range	Low	High
Vfreq	1.00	0.00	1.00	0.59	0.00	0.59	0.18	0.00	0.18	0.00	0.00	0.00
Vwidth	1.00	0.00	1.00	0.59	0.00	0.59	0.18	0.00	0.18	0.00	0.00	0.00
Vrough	0.29	0.71	1.00	0.26	0.35	0.61	0.16	0.07	0.23	0.00	0.00	0.00
Vslope	0.29	0.71	1.00	0.26	0.35	0.61	0.16	0.07	0.23	0.00	0.00	0.00

Figure 7-2. Example sensitivity analysis produced with the spreadsheet programs available through the EL home page on the World Wide Web (Continued)

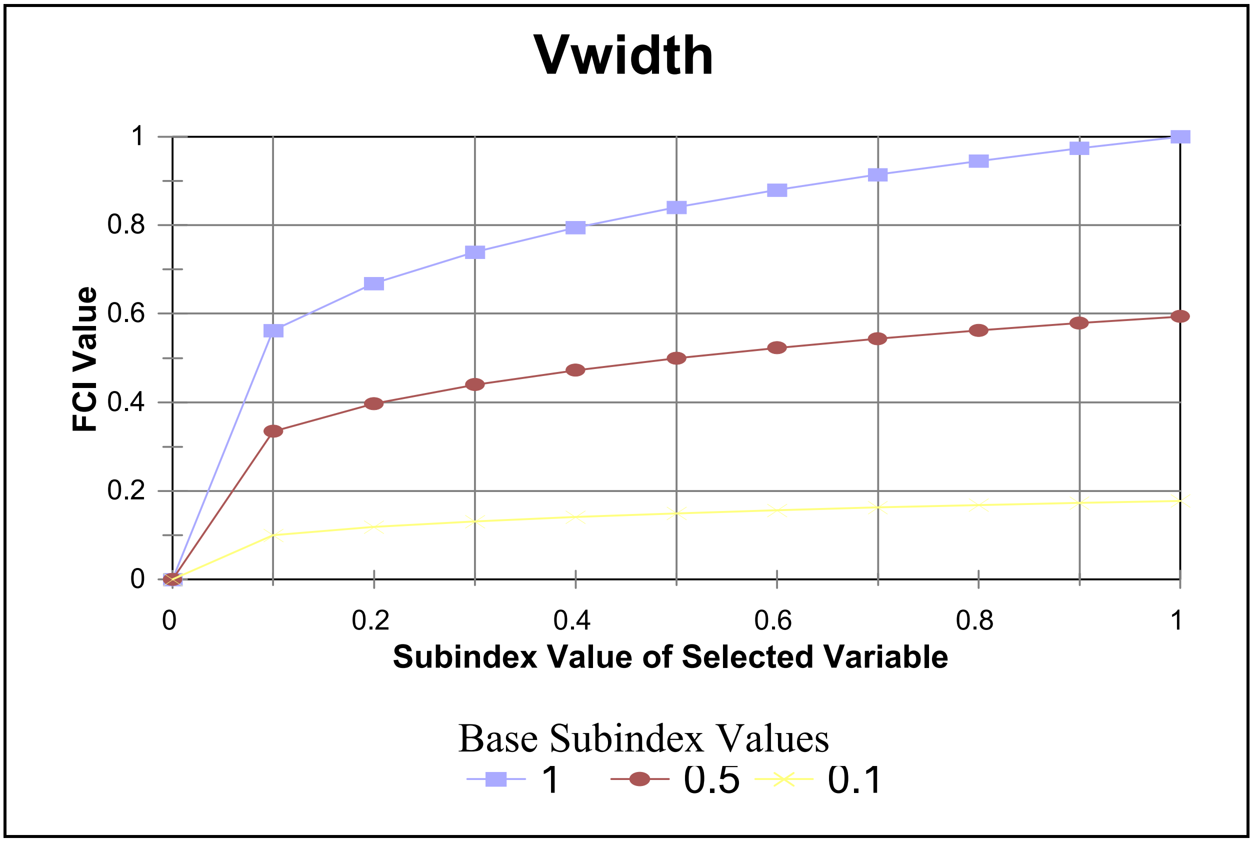

Figure 7-2. (Concluded).

graphical output shown in Figure 7-2 for the variable V_{WIDTH}. The spreadsheet program will produce sensitivity plots for any variable the user requests.

Care must be taken in using the generic spreadsheet programs that inappropriate values of the subindices for certain variables are identified and their effects discounted. For example, the programs automatically increment the subindex for the target variable from 0 to 1. However, the model may specify a different potential range for certain variables. In the western Kentucky, low-gradient riverine model (Ainslie et al. 1999), for instance, V_{SLOPE} takes values only from 0.1 to 1.0; zero is not an appropriate value. Therefore, to investigate the effect of V_{SLOPE} on model outcome, only subindex values from 0.1 to 1 should be considered.

Applying the model to sample data sets

As mentioned previously, a sensitivity analysis that starts at the subindex level cannot verify that the model will respond appropriately to actual values of the field measurements. This can be done only by inputting the actual measures for each variable and examining both the resulting subindices *and* FCI. Appropriate data sets for such an analysis may already be available for wetland sites used in the calibration phase, and additional sites representing a range of conditions can be sampled specifically for this purpose. See Chapter 5 for guidance on collecting and managing reference data. Another option is to generate hypothetical data based on the A-Team's understanding of realistic values for each field variable in the reference domain.

Application of the model to data from a small number (e.g., 10 to 20) of wetland sites can readily be done by hand calculation. Larger applications can be facilitated by programming the complete model (including field measure-to-subindex transformations) into a spreadsheet and inserting either real or hypothetical field data into the appropriate cells. An example spreadsheet is available for downloading through the HGM Web site (http://www.wes.army.mil/el/wetlands/hgmhp.html). Changing variable names, subindex transformations, and the aggregation equation can adapt this spreadsheet for use with any model. Another option, if software is available, is to use a statistical package with programming capabilities to write the model and run the test data.

Figure 7-3 shows an example application of the Temporary Storage of Surface Water model from the western Kentucky, low-gradient riverine guidebook (Ainslie et al. 1999), programmed and run with Statistical Analysis System software (SAS Institute, Inc. 1988). The input data set consists of the actual field measurements from 15 different wetlands. The program first reads the data, then calculates subindex values for each variable based on the graphs given in the guidebook. These subindices are then combined to determine FCI, using the equation given in the guidebook, and results are printed in tabular form.

The output (Figure 7-3) shows that the model produces subindices and FCI values in the appropriate range (i.e., 0 to 1). The A-Team should next determine whether FCI values appear reasonable in light of team members' professional judgment and experience with these particular wetlands. One result of the analysis is that none of these wetlands achieved a FCI score greater than 0.87. If the sample included wetlands judged to meet reference standards, then the model may be scoring these wetlands too low. Furthermore, with the exception of two sites that never flood, no wetland scored lower than 0.26. The A-Team may need to revisit model calibration (Chapter 6) or make other modifications based on the team's best professional judgment.

Checking for correlations among variables

A sensitivity analysis can help to identify variables that have little influence on model outcome and thus could be eliminated to reduce sampling effort and improve the responsiveness of the model to changes in the remaining variables. Another approach that can help to simplify a complex model is to analyze the reference wetland data set for correlations that may indicate redundancies among variables. If two variables are highly correlated, it may be possible to eliminate one without significant loss of information.

```
* Example application of the HGM assessment model for the "Temporary      ;
* Storage of Surface Water" function of the western Kentucky              ;
* low-gradient riverine guidebook, programmed in SAS.                     ;
*                                                                         ;
* Lines preceded by a '*' are comments and do not affect the running      ;
* of the program.                                                         ;
*                                                                         ;
* First, enter the field measurements for each variable, where:           ;
*   SITE = site identifier                                                ;
*   RECUR = estimated flood recurrence interval (years)                   ;
*        (Code as 0 if stream does not flood)                             ;
*   GRADIENT = feet of elevation change per 5280 feet (%)                 ;
*   MANNING = Manning's n                                                 ;
*   RATIO = ratio of channel width to floodplain width                    ;
*                                                                         ;
DATA A;
 INPUT     SITE      RECUR     GRADIENT     MANNING     RATIO;
 CARDS;
            1          5         .02          .05         25
            2          0         .05          .05         10
            3          8         .12          .12         15
            4          2         .08          .08         75
            5          1         .01          .02         22
            6         18         .05          .05          4
            7          6         .09          .10          5
            8         10         .15          .03         40
            9          0         .02          .11         20
           10          1         .01          .04          9
           11          2         .05          .06         12
           12          5         .01          .09         18
           13         12         .08          .03         80
           14          8         .20          .17         12
           15          3         .07          .13          5
;
DATA B;
 SET A;
* Conversion of RECUR to a subindex (Vfreq):                              ;
IF (RECUR >= 1) AND (RECUR <= 2) THEN Vfreq = 1;
IF (RECUR > 2) AND (RECUR < 14) THEN Vfreq = (-.075 * RECUR) + 1.15;
IF RECUR >= 14 THEN Vfreq = 0.1;
IF RECUR = 0 THEN Vfreq = 0;
*                                                                         ;
* Conversion of GRADIENT to a subindex (Vslope):                          ;
IF GRADIENT <=.05 THEN Vslope = 1;
IF (GRADIENT >.05) AND (GRADIENT <.23) THEN Vslope = (-5 * GRADIENT) + 1.25;
IF GRADIENT >=.23 THEN Vslope = 0.1;
```

Figure 7-3. Example application of the Temporary Storage of Surface Water model from Ainslie et al. (1999) programmed and run with SAS software (Continued)

```
* Conversions of MANNING to a subindex (Vrough):                          ;
IF (MANNING >=.11) AND (MANNING <=.15) THEN Vrough = 1.0;
IF (MANNING <.11) AND (MANNING >.03) THEN Vrough = (11.25 * MANNING) -.2375;
IF (MANNING >.15) AND (MANNING <.19) THEN Vrough = (-20 * MANNING) + 4;
IF MANNING <=.03 THEN Vrough = 0.1;
IF MANNING >=.19 THEN Vrough = 0.2;
*                                                                         ;
* Conversions of RATIO to a subindex (Vwidth):                            ;
IF RATIO <= 10 THEN Vwidth = 0.1;
IF (RATIO > 10) AND (RATIO < 70) THEN Vwidth = (.015 * RATIO) - .05;
IF RATIO >= 70 THEN Vwidth = 1.0;
*                                                                         ;
* Calculate FCI according to the equation given in the model:             ;
FCI = (((Vfreq * Vwidth)**0.5) * (Vrough + Vslope)/2)**0.5;
*                                                                         ;
* Print the results:                                                      ;
PROC PRINT;
  VAR SITE RECUR Vfreq GRADIENT Vslope MANNING Vrough RATIO Vwidth FCI;
RUN;
```

The following output was produced by the SAS run:

The SAS System   14:02 Thursday, November 6, 1997   1

OBS	SITE	RECUR	VFREQ	GRADIENT	VSLOPE	MANNING	VROUGH	RATIO	VWIDTH	FCI
1	1	5	0.775	0.02	1.00	0.05	0.3250	25	0.325	0.58
2	2	0	0.000	0.05	1.00	0.05	0.3250	10	0.100	0.00
3	3	8	0.550	0.12	0.65	0.12	1.0000	15	0.175	0.51
4	4	2	1.000	0.08	0.85	0.08	0.6625	75	1.000	0.87
5	5	1	1.000	0.01	1.00	0.02	0.1000	22	0.280	0.54
6	6	18	0.100	0.05	1.00	0.05	0.3250	4	0.100	0.26
7	7	6	0.700	0.09	0.80	0.10	0.8875	5	0.100	0.47
8	8	10	0.400	0.15	0.50	0.03	0.1000	40	0.550	0.38
9	9	0	0.000	0.02	1.00	0.11	1.0000	20	0.250	0.00
10	10	1	1.000	0.01	1.00	0.04	0.2125	9	0.100	0.44
11	11	2	1.000	0.05	1.00	0.06	0.4375	12	0.130	0.51
12	12	5	0.775	0.01	1.00	0.09	0.7750	18	0.220	0.61
13	13	12	0.250	0.08	0.85	0.03	0.1000	80	1.000	0.49
14	14	8	0.550	0.20	0.25	0.17	0.6000	12	0.130	0.34
15	15	3	0.925	0.07	0.90	0.13	1.0000	5	0.100	0.54

Figure 7-3. (Concluded).

The simplest way to evaluate relationships among variables in the reference data set is to use a statistical program to calculate a correlation matrix. Separate correlation matrices should be calculated for field measurements of each variable and for subindices of each variable. Correlation matrices for the variables shown in Figure 7-3 are given in Figure 7-4. The upper tabulation shows correlations among the measurements, and the lower tabulation gives correlations among the subindices. For each pair of numbers, the upper is the Pearson correlation coefficient r and the lower is the associated probability or significance level under the null hypothesis that $r = 0$. In the example, none of the correlations is significant at $\alpha = 0.05$ and the coefficients themselves are relatively small (i.e., maximum $|r| = 0.48$, or $r^2 = 0.23$). As a rule of thumb, one need not be concerned about redundancies between variables until the coefficient of determination (r^2) exceeds 0.50 (or $|r| > 0.70$), indicating that more than 50 percent of the variation in one measurement can be accounted for by changes in the other; r^2 values exceeding 0.80 ($|r| \geq 0.90$) indicate substantial redundancy between two measures.

If two variables are highly correlated, the A-Team should consider simplifying the sampling protocol by eliminating one of the variables from the model. Factors to consider in deciding which of the two variables to keep include (a) ease of making the measurement, (b) accuracy and precision of the measurement, and (c) relevance of the variable to the anticipated wetland impacts in the region.

Field Testing the Model

Field testing helps to ensure that the model can be applied quickly and efficiently by typical users, and that results are consistent and reproducible, at least within limits acceptable to the A-Team. Again, model accuracy is not an issue here (see the section "Validating the Model"). There are no firm guidelines concerning how long it should take to apply an assessment model to a typical field site, nor how consistent results must be from one investigator to the next. Both depend upon the user's constraints and expectations. A-Teams should establish and document realistic goals for time and repeatability in advance of any field testing. For routine regulatory purposes, application of the set of assessment models for all the functions performed by a wetland of a particular regional subclass should probably take no more than a few hours. Requirements for consistency depend upon the intended use of the model. A model that is used to guide multimillion dollar land use decisions should be tested to a higher standard than one intended solely for routine wetland management or advanced identification projects.

An important issue in model consistency is the inherent variability of many quantitative measures across a wetland site and the statistical considerations of sample size and sampling design. Sampling procedures recommended in a Regional Guidebook should be based in part on analysis of data from reference wetlands. Recommended sample sizes (e.g., number of plots or transects) are a trade-off between the desire for a rapid assessment and the need for confidence in the estimates of each variable and FCI. Statistical issues in sampling design are considered in Chapter 5.

The SAS System   14:02 Thursday, November 6, 1997

Correlation Analysis

4 'VAR' Variables: RECUR GRADIENT   MANNING   RATIO

Pearson Correlation Coefficients / Prob > |R| under Ho: Rho=0 / N = 15

	RECUR	GRADIENT	MANNING	RATIO
RECUR	1.00000 0.0	0.41487 0.1241	-0.06177 0.8269	0.12475 0.6578
GRADIENT	0.41487 0.1241	1.00000 0.0	0.48391 0.0676	0.12869 0.6476
MANNING	-0.06177 0.8269	0.48391 0.0676	1.00000 0.0	-0.31890 0.2467
RATIO	0.12475 0.6578	0.12869 0.6476	-0.31890 0.2467	1.00000 0.0

4 'VAR' Variables: VFREQ   VSLOPE   VROUGH   VWIDTH

Pearson Correlation Coefficients / Prob > |R| under Ho: Rho=0 / N = 15

	VFREQ	VSLOPE	VROUGH	VWIDTH
VFREQ	1.00000 0.0	0.08392 0.7662	0.03643 0.8974	-0.00061 0.9983
VSLOPE	0.08392 0.7662	1.00000 0.0	-0.08350 0.7674	-0.09806 0.7281
VROUGH	0.03643 0.8974	-0.08350 0.7674	1.00000 0.0	-0.28959 0.2951
VWIDTH	-0.00061 0.9983	-0.09806 0.7281	-0.28959 0.2951	1.00000 0.0

Figure 7-4. Correlation matrices for field measurements (upper tabulation) and subindices (lower tabulation) for the data shown in Figure 7-3

This section describes a generic procedure for model field testing. The procedure is adaptable to different levels of effort in data gathering and analysis, depending on the needs of and constraints upon the A-Team. A relatively simple test might involve only a small number of participants (e.g., 6 to 10) and a few field sites. However, conclusions drawn from such a limited sample would be questionable and could not give the A-Team much confidence in the repeatability of model scores across investigators. Larger samples of test participants (e.g., 25 or more) are needed to determine the distribution of FCI scores and to give the A-Team more confidence that different investigators assessing the same site will obtain similar results.

A generic procedure

Table 7-1 lists the steps involved in a field test of a draft assessment model. This procedure can be used to test the model for a single function or the set of functional models performed by a wetland subclass. Models for different functions often use some of the same variables; therefore, a realistic evaluation of the amount of time required to apply the set of models for that subclass is possible only if the models for all functions are applied at once.

Table 7-1
Sequence of Steps Involved in a Generic Field Test of a Draft HGM Assessment Model

Step	Description
1	Identify a number of individuals to serve as field testers. The larger the sample of testers, the more reliable the conclusions about the distribution of model scores.
2	Select at least three to five wetland field sites representing a range of conditions relative to reference standards.
3	Provide the draft guidebook (including models, instructions, and data forms) and background site information to testers in advance of site visits.
4	Schedule site visits by each tester independently, if possible. In any case, testers should not be influenced by other test participants. Consider scheduling two or more rounds of tests to evaluate seasonal or annual bias.
5	Ask testers to record the amount of time required to apply the model at each field site and, after completion of all field visits, to provide a written critique of the model instructions, sampling procedures, and calculations.
6	Combine field results from all testers. Evaluate consistency of FCI scores across testers for each wetland function considered.
7	If model output is inconsistent, modify the model, instructions, or sampling recommendations to reduce variability. If necessary, schedule a new field test using some of the same and some different participants.

Note: The purpose of the test is to determine time requirements for applying the model and to evaluate consistency of results across different investigators. Model accuracy is not considered. See text for details.

The first step in field testing the models is to identify a number of individuals willing to serve as testers. The A-Team should choose people who were not involved in the development of the models, sampling protocols, or the instructions for their use. It is important to select individuals whose training and experience are similar to those of anticipated end users of the models (e.g., regulatory personnel, private consultants, resource managers). All participants should have experience with basic methods for sampling environmental characteristics.

Next, select a manageable number of wetland field sites (at least three to five sites is suggested) of the appropriate regional subclass within the intended reference domain. Include at least one site that represents reference standard conditions and two or more that deviate from reference standard. Some of the same reference wetland sites used for model calibration may be adequate for this purpose; it is not necessary to select new sites. To test consistency of model output, it is more important to maximize the number of testers than it is to increase the number of sites. A field test involving 20 people and 3 field sites is likely to provide more useful data than one involving only 6 people and 10 sites.

Each model tester should be provided in advance with the models, field data forms, sampling protocols, and detailed instructions for their use. In addition, background information on the field sites should be provided, including topographic maps, soil survey information, National Wetlands Inventory maps, hydrology data, and any other office data required by the models. Testers should be thoroughly familiar with the instructions for using the models before they go to the field.

It is important for each individual tester to provide an independent determination of FCI for a site, unswayed by other participants in the test. The preferred option is to schedule separate site visits by each model tester, if possible. If separate visits are not practical, take steps to ensure that participants do not interact, cooperate, or interfere with each other during the tests.

Two potential goals of field testing are to evaluate (a) the clarity of instructions for applying the guidebook by assessing the consistency of results across different individuals and (b) seasonal or annual variations in FCI scores produced by the model. All draft assessment models should be evaluated for investigator consistency (goal 1). To do this, it is suggested that all field testers be scheduled for site visits within a 1- to 2-week period to minimize the influence of temporal changes in site conditions on FCI scores. In addition, any model that contains variables whose interpretation might change seasonally (e.g., spring versus summer) or annually (e.g., wet vs. dry years) should also be evaluated for temporal consistency (goal 2). This can be done by scheduling two or more rounds of field tests during different seasons or years (see “Evaluating temporal consistency”).

Upon arrival at a field site, testers should be oriented relative to site maps and important landmarks, made aware of the boundaries of the wetland assessment area, provided with any necessary tools, and then asked to perform the assessment. Each tester should record the amount of time required to gather field data at each site, and should use his or her data to determine subindices for each variable and FCI values for each function. After completion of sampling and data analysis at all field sites, testers should be asked to provide written comments addressing the clarity, completeness, and "user friendliness" of the instructions for applying the models, sampling procedures, and calculations. A form such as that shown in Figure 7-5 may be used for the testers' comments.

FCI scores for each function at each field site are then compiled and compared to evaluate consistency in scoring by different testers. As mentioned previously, there are no established standards for consistency of model outputs across investigators and the desired precision may vary with the goals of the application (e.g., general resource inventories versus high-value impact analyses). Therefore, the A-Team should establish goals for investigator consistency in advance of field testing. For most regulatory uses, including wetland impact assessments, project alternatives analyses, and calculation of mitigation requirements, the following test goal is suggested: 90 percent of users who apply a model in the same assessment area should produce a FCI score that is within 0.15 of the median score for all users combined.

As an example, Table 7-2 presents the results of a simple field test involving six participants who were asked to apply a set of five functional assessment models to a series of sites. Results for only one field site are shown. Due to the small number of testers involved, analysis of these data is necessarily subjective and the application of standards must be flexible. FCI scores for Functions 1, 3, and 5 clearly meet the goal in that all six test participants achieved scores within 0.15 of the median score for each function. Results for Function 4 are very consistent (5 of 6, or 83 percent, achieved the same score) with the exception of that obtained by David Moran. Examination of the written comments provided by the testers are valuable in reconciling outlying scores. In this case, the low score by David Moran may reflect his confusion over some part of the instructions that could be corrected easily. The fact that other testers gave consistent scores may indicate that the instructions and model documentation are basically sound.

Model consistency must be evaluated across all field sites involved in the test, particularly those representing moderate departures from reference standard conditions. Inconsistencies may be more obvious and informative at sites having intermediate levels of function than at sites representing the extremes. For example, the perfect consistency among users of the model for Function 5 (Table 7-2) at that site may be due to some obvious limitation (e.g., the function requires surface flow and the site never floods); this does not mean that model outcome would be consistent among users on a site that does flood.

Scores for Function 2 (Table 7-2) are highly variable. The model for this function clearly fails to meet the stated goal for investigator consistency.

HGM Assessment Model Field Test -- Tester's Evaluation Form

Tester's Name: ______________________________ Model: ________________________________
Phone: ______________________________ Date: ________________________________
E-Mail: ______________________________

Time Required to Apply Model(s):

Field Site 1:	Start time: _________	Completion time: _________	Total time elapsed: _________
Field Site 2:	Start time: _________	Completion time: _________	Total time elapsed: _________
Field Site 3:	Start time: _________	Completion time: _________	Total time elapsed: _________
Field Site 4:	Start time: _________	Completion time: _________	Total time elapsed: _________
Field Site 5:	Start time: _________	Completion time: _________	Total time elapsed: _________

To apply the model(s), did you need any documents or tools that were not available? Please list:

Did application of the model(s) require particular training or experience that you lacked? Please list:

Were the written instructions complete? If not, identity gaps that need to be corrected:

Were the instructions clearly written and easy to follow? Identify specific problems or ambiguities:

Describe any general problems you had in determining subindex levels for each variable.

Describe any general problems you encountered with calculation of FCI values.

Figure 7-5. Example field tester's evaluation form (Continued)

For each variable listed below, give your opinion as to (1) the clarity of the instructions for measuring that variable in the field, (2) ease of making the field measurement, and (3) whether conversion of the measure to a subindex was clear and straightforward. Use the following scale for your response:

1-strongly disagree, 2-disagree, 3-no opinion, 4-agree, 5-strongly agree

Variable	**Sampling instructions were clear**	**Field measurement was easy**	**Conversion to subindex was straightforward**
V_A			
V_B			
V_C			
V_D			
V_E			
etc.			

For each function listed below, give your opinion (1) whether calculation of the FCI was clear, and (2) whether the FCI agreed with your subjective opinion of the quality of the site(s) for that function. Explain any differences of opinion. Use the scale given above for your responses.

Function	**FCI calculation was clear**	**FCI agreed with my subjective judgement**
Function #1		
Function #2		
Function #3		
etc.		

Do you think that the instructions for using this model in the field are ready for publication and distribution? If not (and not covered above), please describe what needs to be done:

Figure 7-5. (Concluded).

Table 7-2
Example Comparison of Field Testers' Results at One Field Site

Tester	FCI Scores				
	Function 1	Function 2	Function 3	Function 4	Function 5
Margaret Diaz	0.3	0.9	1.0	0.8	0.0
John Engles	0.25	0.6	1.0	0.8	0.0
Ellen Frances	0.3	0.8	0.9	0.8	0.0
JoAnne King	0.2	0.3	1.0	0.8	0.0
David Moran	0.25	0.6	1.0	0.4	0.0
Cindy Wong	0.3	0.5	1.0	0.8	0.0
Scoring Summary:					
Min./Max.	0.2 - 0.3	0.3 - 0.9	0.9 - 1.0	0.4 - 0.8	0.0 - 0.0
Median	0.275	0.6	1.0	0.8	0.0

Figure 7-6 shows the distribution of FCI scores for a different field test involving a larger number of participants (*n* = 30) and models for two functions. Again, results for only one field site are shown. The larger sample size provides more information about the distribution of FCI scores than did the previous example. When the same goal for investigator consistency is applied, the model for Function 1 passes the test (i.e., 28 of 30 FCI scores, or 93 percent, fall within 0.15 of the median score for all test participants). Scores for Function 2, however, are too variable. Only 47 percent (14 of 30) of FCI scores fall within the desired range (Figure 7-6).

There may be several reasons why a model would fail to meet goals for investigator consistency, including (a) unclear definitions of model variables, (b) use of low-resolution or error-prone sampling methods, (c) unclear instructions for data gathering, and (d) investigator errors in calculating subindex and FCI values. In addition, if an assessment area is large or heterogeneous, sample sizes recommended in the Regional Guidebook may not be large enough to achieve adequate precision in the estimates of quantitative variables (see Chapter 5). This problem can be corrected by requiring larger samples (e.g., more plots) at the expense of application speed.

Sometimes the problems with model consistency can be traced to only one or two variables. Written comments provided by model testers are valuable in identifying such problems and providing suggestions for model improvement. Another way to identify problem variables is to plot histograms of subindex scores, similar to the plots of FCI values shown in Figure 7-6. Model revisions should aim at reducing the variability in individual subindex scores.

Assessment models that undergo extensive changes as a result of a field test should be tested again to determine whether the consistency of model scores across investigators has improved. For a repeat field test, some of the same and some different participants should be used.

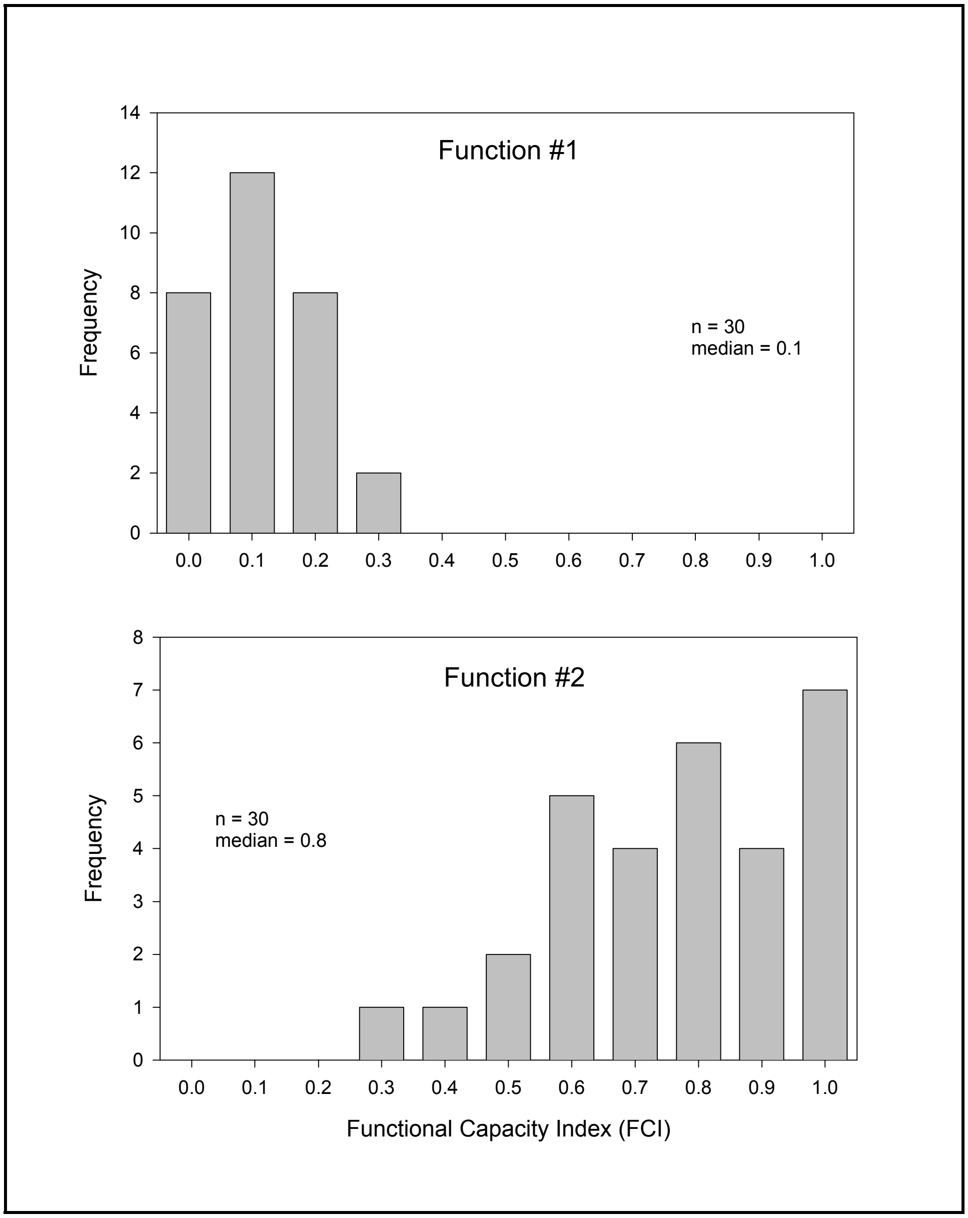

Figure 7-6. Results of field tests of assessment models for two functions at one field site. Histograms of FCI scores were based on independent determinations by 30 participants in the test. See text for explanation

Evaluating temporal consistency

Most HGM assessment models are designed to be used throughout the year, at least when weather conditions are adequate for sampling (i.e., snow is not too deep and soils are unfrozen). Some models may contain alternative variables to use at particular times (e.g., when surface water is present versus absent). However, the potential of a wetland to perform certain functions does not change seasonally or annually in undisturbed situations. Therefore, a model applied to a particular wetland should give the same score regardless of when the investigation is done.

Any model that contains variables that may change seasonally or annually should be field tested for temporal consistency. This includes models whose variables may be more difficult to evaluate during certain periods, such as during dry seasons or years. Temporal consistency is evaluated by applying the model at different times and comparing the results.

Table 7-3 shows the results obtained by 12 field testers who applied the model for one function at one wetland site during spring and again in late summer. This example used the Mann-Whitney test, a nonparametric analogue to the t-test (Zar 1984), to determine whether distributions of FCI scores differed between sampling dates. The lack of a significant difference in FCI scores indicates that the model gives results that are consistent through the year.

In some cases, it may be possible to modify the model to improve temporal consistency without reducing potential model accuracy by emphasizing only the most stable environmental features or measurements. For example, a wildlife model might require an estimate of acorn availability as one variable affecting winter food supplies. A direct measurement of the abundance of fallen acorns would vary seasonally, especially in areas affected by flooding. A surrogate variable, such as density of acorn-producing trees, is temporally more stable and could be measured at any time of year.

In other cases, critical environmental measurements may be impossible to take at certain times of year (e.g., pH of surface water or maximum flow velocity when wetlands are dry). One option is to include less reliable indicator variables as alternatives if the assessment must be done at an inappropriate time. Model documentation should include the results of a consistency test (e.g., Table 7-3) and should state that model scores may be less reliable if an indicator variable must be substituted for the preferred measurement.

Validating the Model

The way to ensure the accuracy and reliability of a model as an index to the magnitude of wetland function is to validate it by comparing its performance against an appropriate standard of comparison (Caswell 1976; Schamberger and O'Neil 1986; Rykiel 1996). For HGM assessment models, that standard is an independent, quantitative measure of function. An assessment model will be useful in the Section 404 process only if (a) it accurately reflects differences in magnitude of function between different wetlands, at least within specified standards of precision, and (b) any change in magnitude of function due to a project results in a proportionate change in the index. These criteria are particularly important if function lost at one wetland (e.g., a project site) is to be replaced at a different wetland (e.g., a mitigation site).

Table 7-3
Field Test for Temporal Consistency of Assessment Model Results for One Function

Tester	FCI Values for Function 1			
	Spring Sampling (25-30 April)		Late Summer Sampling (1-15 August)	
	FCI	Rank	FCI	Rank
Roy Banks	0.4	6	0.3	2
Scott Barber	0.5	13.5	0.5	13.5
Linda Hammond	0.5	13.5	0.4	6
Mellissa Hrosovski	--	--	0.5	13.5
Alonzo Jackson	0.6	20	0.5	13.5
Margaret Johnson	0.4	6	0.3	2
Otis Kenworthy	0.5	13.5	--	--
John Kindhart	--	--	0.4	6
Mercedes Lebeau	0.3	2	0.5	13.5
José Lopez	0.5	13.5	0.6	20
Deborah Patterson	0.4	6	0.6	20
Kathy Rittenhauer	0.5	13.5	0.5	13.5

Example two-tailed Mann-Whitney test with tied ranks (Zar 1984):

Sample sizes: $n_1 = 10$
$n_2 = 11$

Sum of ranks: $R_1 = 107.5$
$R_2 = 123.5$

Test statistic: $U = n_1 n_2 + [n_1(n_1 + 1)]/2 - R_1$

$U = (10)(11) + (10)(11)/2 - 107.5 = 57.5$

The critical value of $U_{10,11}$ at $\alpha = 0.10$ for a two-tailed test is 79. Therefore, it was concluded that there was no significant difference between FCI scores determined in April and August.

Note: FCI scores were derived from repeat sampling of the same wetland in spring and late summer.

Model validation can be an expensive and time-consuming proposition. It probably will be years before a significant number of models have been subjected to rigorous validation. However, this does not mean that use of assessment models in the regulatory arena should wait until validation can be done. The need for regulators to assess the potential impacts of a project to wetland functions is already here (Smith 1993). Assessment models currently under development represent the best available technical input to those decisions, whether or not the models have been validated.

Why validate?

Validation ensures that project-related changes in wetland function are reflected accurately by changes in both the direction and magnitude of the index. This in turn ensures that wetland impacts and mitigation credits will be estimated comparably, and that there will be no unintended gain or loss in wetland function due to a project. Model validation has additional practical advantages to both model developers and end users, including the ability to:

a. *Maintain and strengthen the scientific foundations of the HGM Approach*. Although assessment models are developed by regional wetland experts familiar with the technical literature, and incorporate data from reference wetlands, no one can predict how well the model will mimic the functioning of a complex wetland ecosystem. A model is a simple abstraction of the complex system; it is a hypothesis that must be tested to determine its worth. Therefore, the scientific method needs to be applied at the end of the model development process as well as at the beginning.

b. *Reduce subjectivity and ambiguity in the definition of wetland functions*. The requirement that models be amenable to validation dictates that functions be defined clearly and quantitatively, leaving no doubt as to the process being modeled and the appropriate independent measure of function against which to test model accuracy.

c. *Reduce individual bias in model development and application*. The A-Team can be unduly influenced by one or more dominant members or by individuals with a particular agenda. The expectation that models will be validated reduces the incentive to "fix" a model.

d. *Provide an objective basis for choosing between alternative models*. In the future, as assessment models proliferate, more than one model may become available for the same wetland function performed by the same regional wetland subclass. Alternative models may be developed by different teams of experts and may make very different predictions about the magnitude of project-related impacts to a function. The only way to determine the "best" model is by validation. In addition, an A-Team may choose to develop more than one version of a model, with the expectation that validation will identify the best model.

e. *Reduce arguments and litigation over reliability of HGM assessment models*. The purpose of the HGM Approach is to provide input into Section 404 permit decisions, which in turn can affect the construction plans and property values of permit applicants. Thus, the reliability of assessment models is likely to become as controversial a topic as wetland delineation has been in the past. Untestable models will simply invite arguments and legal challenges.

Who validates?

It seems logical that model developers would have the most interest in pursuing model validation. However, model validation is beyond the immediate mandate and financial resources of most A-Teams, which consist largely of volunteers. End users (e.g., regulators, consultants, developers, wetland managers) are also logical candidates for performing validation work; the incentive to initiate such studies may depend upon the economic value of the intended application. Third parties (e.g., university researchers and their graduate students) are also likely providers of validation work.

Whether or not the A-Team is directly involved in model validation, it is their responsibility to ensure that models are *amenable* to validation. In particular, it is critical for the A-Team to specify an independent, quantitative measure for each function modeled (see Chapter 4 for guidance on defining wetland functions.)

Approaches to model validation

There are two basic approaches to validating assessment models. The first involves experimental manipulation of site characteristics at one or more wetland sites to see whether the model is able to predict observed changes in the magnitude of function. For example, a model for Particulate Retention may predict that the sediment-trapping capacity of a floodplain wetland will be reduced by 30 percent if all large trees are removed. A test of the model might consist of measuring sediment accretion for a period of time under existing conditions, then harvesting all large trees and measuring the change in accretion rates. This approach may provide the truest test of model performance (Schamberger and O'Neil 1986), given that the primary use of HGM assessment models is to predict changes in wetland function due to project-related disturbances. However, manipulative experiments may be difficult to accomplish due to the time required and the need for wetland sites that can be altered at will.

The second approach to model validation is to evaluate the correlation between model output and actual measurements of the magnitude of function at a series of reference wetland sites selected to represent a range of capacities for the function of interest. This approach does not involve manipulation of any wetland, although the amount of time and effort required to accomplish the test will depend upon the difficulty of directly measuring the magnitude of function at each site. The following sections will focus on this second approach to model validation because it is likely to be more practical and more often used. Correlation of model output against a measured standard of comparison has been used extensively to validate Habitat Suitability Index (HSI) models developed for use with the Habitat Evaluation Procedures (HEP) (U.S. Fish and Wildlife Service 1980, 1981). Some examples relevant to testing of HGM assessment models include Lancia et al. (1982), Cook and Irwin (1985), O'Neil et al. (1988), O'Neil (1993), and Adamus (1995). Terrell and Carpenter (1997) summarized dozens of published and unpublished HSI model tests.

Independent measures of function

The appropriate standard of comparison for HGM assessment models is an independently derived quantitative measure of the function of interest against which model performance can be evaluated. For example, if one wished to validate a model for Particulate Retention in depressional wetlands, an appropriate standard of comparison might be an estimate of the amount of sediment retained by a wetland per unit area per year. For a model of Dynamic Surface Water Storage, the appropriate standard might be a measure of the volume of floodwater retained over a specified time period. A Nutrient Transformation model might be validated against estimates of the number of kilograms of nitrate transformed per unit area per year. A model for Maintenance of Wildlife Communities could be tested by comparing its output against estimates of the number of species of breeding vertebrates in a series of wetlands. The independent measure of function appropriate to each model should be stated in the Regional Guidebook as part of function definition. Additional examples of quantitative measures of function are given in Chapter 4.

One reason that it is important for the Regional Guidebook to specify an independent measure of function for each model is that different measures may not be strongly correlated with one another. Therefore, use of the wrong standard can result in rejection of a model that, in fact, may be valid for its intended purpose. For example, the developers of a Wildlife Community model may have intended that the model predict changes in the number of breeding bird species using a wetland. Thus the model would contain variables relevant to birds. The abundance of frogs and salamanders in the same wetlands might be influenced by quite different habitat features, and may not be related to the diversity of birds at all. Therefore, it would be inappropriate to test the draft model against an estimate of amphibian diversity.

Because of the difficulties involved in measuring wetland functions directly, it may be useful for some purposes to perform a preliminary validation based on subjective estimates of function provided by experts at a series of reference wetlands. For example, the A-Team might use the results of such a "prevalidation" to perform a final calibration of model variables (see Chapter 6) or to select the most appropriate aggregation equation for the operational draft model. Subjective estimates of function should be obtained from experts not previously involved in assessment model development. The experts should be taken to each site as a group, asked to rate the magnitude of function on a relative scale (e.g., from 0 to 1.0), and asked to provide written documentation of the reasons for their scoring. They should first score the wetland independently, then confer and arrive at a consensus score. Consensus scores can then be used in place of independent measures of function in the procedures described in the following section. Prevalidation based on expert opinions is clearly not a substitute for validation based on independent measurements of function; however, it may be useful in the development of an operational draft model.

Expected relationship between FCI and the independent measure of function

The expected relationship between FCI and an independent measure of function must be considered in any model validation study. In the HGM Approach, the FCI is an index expressed on a ratio scale ranging from 0 to 1 (see Chapter 4). There are two important features of a ratio scale (Zar 1984). First, the interval size between adjacent units is constant. Thus, a change in FCI from 0.2 to 0.3 represents the same magnitude of change as one from 0.8 to 0.9. Second, there is a physically significant zero point on the scale. In the HGM Approach, zero FCI represents the condition in which the wetland function or process does not occur (Smith et al. 1995).

Direct measures of wetland functions are also characterized by ratio scales. Examples include counts of items, lengths, weights, volumes, rates, and units of time (Zar 1984). The HGM Approach assumes that there is a one-to-one relationship between FCI and the magnitude of function. Consequently, the expected relationship between FCI and an independent measure of function for all functions is *linear* (Figure 7-7, Graph A). Generally, the absence of a wetland function or process at a site will be indicated by a zero for the independent measure of function (e.g., no wetland wildlife present, no sediment trapped, no organic carbon exported). Therefore, a plot of FCI versus the independent measure of function usually should pass through the origin (Ott 1978). The level of function that corresponds to FCI = 1.0 (denoted as X_{RS} in Figure 7-7, Graph A) will vary with wetland subclass, region, and other factors. The numerical value of X_{RS} is determined by the actual magnitude of function in reference standard wetlands and is estimated during model calibration (see Chapter 6). Therefore, X_{RS} is not a single number but is a range of values dictated by the range of function encountered among different reference standard wetlands. For example, four floodplain wetlands deemed to be reference standards may export organic carbon at rates ranging from 21 to 35 kg/ha/year. This range would constitute X_{RS} and would correspond with FCI = 1.0 (Figure 7-7, Graph B).

For some functions, there may be no disadvantages to even higher levels of function and, therefore, no decline in FCI. Say, for example, that the number of breeding forest-interior bird species at reference standard sites ranged from 13 to 16; this range would represent X_{RS} in an assessment model designed to predict species richness of forest-interior birds. However, another site not considered to be a reference standard may contain 18 species. For this function, the A-Team would probably design the model to give sites having unusually high levels of function (i.e., bird richness) an FCI of 1.0 (e.g., line *a* in Figure 7-7, Graph B). On the other hand, certain functions, when they occur at unusually high levels, may not be sustainable and may contribute to wetland degradation or destruction. Brinson (1995) used the example of increased rate of sediment transport into a wetland due to clearing of surrounding upland forests. Unsustainable levels of sediment input and retention in a wetland should result in a decline in FCI (e.g., line *b* in Figure 7-7, Graph B), although the slope of this line may be arbitrary or based on assumptions made by the A-Team.

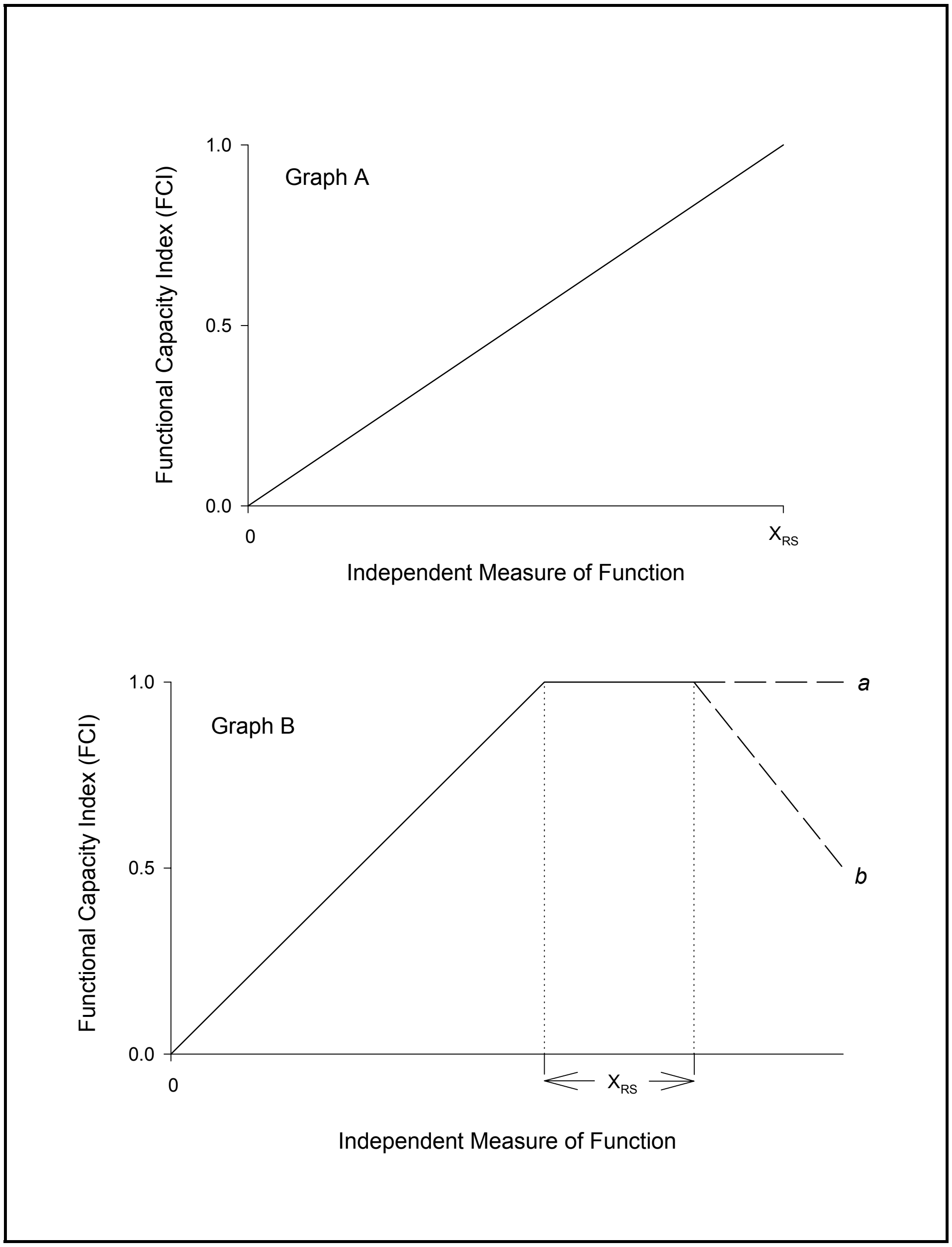

Figure 7-7. Expected relationship between the modeled FCI and a direct measurement of that function. See text for details

There are two main reasons why the relationship between FCI and an independent measure of function may differ from the expected relationships depicted in Figure 7-7. First, there may be some problem with the draft model (considered in more detail in the following sections), or second, there may be error and/or bias in the independent measurement of the function. The basic design of a model validation study is to identify a series of wetland sites, apply the model to each site to estimate FCIs for a particular function, independently measure the magnitude of function, compare FCI and the independent measure of function across all sites, and, if needed, modify the draft model to bring FCI scores and independent measures into better agreement. This approach assumes that the independent measure is the "correct" measure of function and that model output should closely match it. However, even direct measures of wetland function are essentially just estimates based on a particular sampling design and measurement technique. Some techniques may be biased (i.e., they may consistently overestimate or underestimate the true level of function), and all direct measures incorporate some level of measurement error that causes "scatter" in the data set. There is little one can do about measurement bias, except to select the most reliable techniques by reviewing published literature and talking with experienced individuals. Measurement error depends on the sampling design and sample sizes used to estimate the magnitude of function; the amount of error can be quantified statistically (e.g., standard error of the mean, confidence limits). It is important to remember that both measurement error and bias in the independent measure of function can reduce the strength of the relationship with FCI.

Testing the whole model or its components

HGM assessment models may have several parts or components, each of which is amenable to testing and validation (Schamberger and O'Neil 1986). At the most basic level, validation studies can target the assumptions underlying the use of specific variables and measures. Many models incorporate variables that are surrogates for or indicators of the actual quantity of interest (Smith et al. 1995). For example, in the absence of a direct measure, the variable Frequency of Overbank Flooding might be evaluated based on indicators of flooding (e.g., presence of wrack lines or silt deposits) or characteristics of the vegetation (e.g., proportion of the dominant plant species in the community that is wetland species). A test to validate this assumption would examine indicators present in areas of known flooding frequency. Similarly, a model for Organic Carbon Export from forested riverine wetlands may use Tree Canopy Cover as a variable under the assumption that canopy cover is directly related to the abundance of organic debris available for export. Validation of this assumption might involve comparisons of canopy cover measurements against the mass of leaves and twigs collected in litter traps within a number of floodplain wetlands. Testing the validity of indicators can be critical to the quality of assessment models because many important variables (e.g., hydrologic and biogeochemical variables) are difficult or impractical to measure directly. The use of indicator variables in models introduces additional variability that can weaken the relationship between model output and actual measurements of wetland function. Careful validation and variable selection can reduce unwanted variability and improve model accuracy.

A second level of validation involves testing the relationship between the measure for each variable and its functional subindex. Like FCI, subindices range from 0 to 1 and are indices to magnitude of function. Therefore, relationships between variables and subindices can be tested by plotting each variable against the independent measure for the function. This relationship should approximate the variable/subindex curve or histogram given in the draft model, except for the effects of other variables in the model (see the section "A generic procedure for model validation" for further details).

Finally, a validation study might target the whole model at once. For example, one might test the accuracy of a model for Wildlife Community Support by first applying it at a number of wetland sites and calculating the FCI for each site. Then, the independent measure of function specified in the model (e.g., combined species richness of breeding terrestrial vertebrates) is measured at each site using appropriate sampling techniques for each component of the vertebrate community (i.e., birds, mammals, reptiles, and amphibians). The combined number of species of vertebrates at each site is calculated, plotted against FCI, and compared with the expected relationship shown in Figure 7-7.

In practice, one should probably start by testing the whole model and then, if needed, examine one or more of the components or underlying assumptions of the model. A model that passes the first test may not need to be tested further for users to have confidence in its predictions. This is no guarantee, however, that all components of the model are necessary or are performing properly. Furthermore, if the overall model does not meet performance expectations, it will be necessary to test each of its parts. As mentioned previously, the purpose of validation is not to reject the model, but to modify the model until its performance meets the goals set by the A-Team. Model validation is an iterative process involving testing, modifying, and retesting until standards for reliability are achieved.

A generic procedure for model validation

The following suggested procedure for validation of HGM assessment models is similar to the method described by O'Neil et al. (1988) for testing and modifying HSI models. It is based on correlations between model output and an independent measure of function. An outline of the procedure is given in Table 7-4.

The first step is to identify a number of reference wetlands from the intended reference domain. At least 10 to 20 sites are recommended. Only two or three sites representing reference standard conditions (i.e., FCI = 1.0) should be included. All others should represent the range of less-than-reference-standard conditions for the function of interest and for each of the variables in the model. It is permissible to use some of the same sites used in model calibration if they meet the guidelines, as long as the calibration step did not already include consideration of actual measures of function at those sites. Otherwise, it will be necessary to select an independent sample of sites to validate the model.

Next, the assessment model is applied at each site and FCI values are calculated (Table 7-4, Step 2). At the same time, variables needed for any alternative versions of the model should also be collected. The purpose of validation is to improve model performance either by modifying the draft model or by replacing

Table 7-4
A Generic Procedure for Validating HGM Assessment Models Based on Correlation of FCI with an Independent Measure of Function

Step	Description
1	Select at least 10 to 20 reference wetlands representing a range of conditions for the function of interest and for each of the variables in the model.
2	Apply the model and calculate FCI for each site. At the same time, collect any variables being considered for alternative versions of the model.
3	Make independent, quantitative measurements of the magnitude of function at each site. Use an accepted sampling method and a design that minimizes bias and measurement error. More than one year of effort may be required to determine average conditions.
4	Based on independent measures of function, reevaluate assumptions made during model development and calibration about reference standard wetlands and the level of function that corresponds with FCI = 1.0.
5	Examine plots and coefficients of determination r^2 of FCI versus the independent measure of function. The expected relationship is linear, as in Figure 7-7, at least for the ascending limb of the graph.
6	Examine plots of the relationships between the measure (x-axis) for each variable in the model and the independent measure of function (y-axis). The plots should resemble the curves or histograms given in the model, except for the effects of other variables on model output.
7	If needed, modify variable measure/subindex relationships, add or drop variables, or adjust the model aggregation equation to improve the correlation between FCI and the independent measure of function. Also test and compare the performance of any alternative versions of the model.
8	If possible, return to Step 1 and initiate a new validation study on the modified model using a different set of field sites.

it with an alternative model having superior performance. Therefore, the investigators should have alternative versions of the model in mind when designing the validation study, and should collect any needed variables at the time each field site is sampled.

The next step is to measure the actual level of function at each site, using the intended independent measure of function stated in the guidebook function definition (Table 7-4, Step 3). Obviously, this step can be difficult and may involve more than one year of effort to determine typical levels of function at each site. The FCI predicted by an assessment model is meant to indicate the normal or average level of function by a wetland; FCI for an undisturbed wetland should not vary appreciably from year to year, unless succession is important to the level of function. However, actual measures of function (e.g., tons of sediment trapped, cubic meters of surface water retained, kilograms of carbon exported, number of breeding vertebrate species detected) do vary annually and more than one year may be needed to determine average conditions.

As mentioned earlier in the section "Expected relationship between FCI and the independent measure of function," investigators should select measurement techniques that are known to be unbiased and use sampling designs that minimize sampling error. This is because any variation in the independent measure of function will affect the strength of the relationship between that measure and FCI. Appropriate measurement techniques can be identified from the literature or by consulting experts. Often specialized equipment or skills are needed, requiring trained and experienced personnel. In addition, the sampling design (e.g., sample size, replication, stratification) and statistical treatment of the data must be carefully planned to minimize error and keep the precision of the measurements within acceptable limits. A measure of precision (e.g., standard error or confidence limits) should accompany each estimate of the independent measure of function.

After values of both FCI and the independent measure of function have been obtained for each wetland site, the first step in data analysis is to reevaluate assumptions made during model development and calibration about the level of function in reference standard wetlands (Table 7-4, step 4). Reference standards are selected based not on one function but on the suite of functions performed by high-quality, relatively undisturbed wetlands in the reference domain (Smith et al. 1995). Now is the time to consider, based on actual measurements of one or more functions of interest, whether wetlands initially selected as reference standards actually deserve that status. The decision is necessarily subjective, but might be based on the measured level of function at a designated reference standard site in relation to the A-Team's a priori opinion of that site. For example, if the A-Team's concept of a reference standard was initially thought to include sites with capacities for Carbon Export in excess of 90 kg/ha/year, then a site with measured carbon export of 75 kg/ha/year may not be an appropriate reference standard site and could be dropped from that status. On the other hand, if other considerations still argue to retain that site among the reference standards, then the implied range of function for reference standard sites (X_{RS}, Figure 7-7, Graph B) must be modified to include sites that export only 75 kg/ha/year. In any case, the draft model's assumed value of X_{RS} should be reevaluated and modified, if necessary, based on actual measures of function at these sites.

The next step in the validation study is to compare FCI values generated by the model against the independently derived measure of function for each wetland site (Table 7-4, step 5). The expected relationship is linear with a y-intercept of 0.0, and with FCI = 1.0 when the independent measure of function equals reference standard X_{RS} (Figure 7-7). Therefore, the strength of the relationship can be evaluated with a linear (Pearson) correlation coefficient r and coefficient of determination r^2. The coefficient of determination is an estimate of the proportion of variability in FCI that is due to its relationship with the independent measure of function (Zar 1984).

Validation should focus mainly on the ascending limb of the relationship between FCI and the independent measure of function (Figure 7-7). This is because the slope of the descending limb, if any, is based mainly on professional judgment of the A-Team, rather than any underlying quantifiable relationship between FCI and the measure of function. In addition, during model validation, modeled FCI values from wetland sites that have measured levels of function within the optimal range (X_{RS} in Figure 7-7, Graph B) should be consolidated and plotted in relation to the lowest value in the X_{RS} range. This procedure eliminates the plateau in the curve, resulting in an expected relationship similar to Figure 7-7, Graph A, and thus makes the relationship more amenable to testing with linear correlation.

Figure 7-8 shows a plot of FCI and an independent measure for a Wildlife Habitat Support function that was intended to reflect the number of species of breeding amphibians present in a wetland. The independent measure was made by counting amphibian species captured during 10 days of trapping in spring using five clusters of pitfall traps (e.g., Block et al. 1994) in each of 10 different wetlands. Ninety-five percent confidence intervals were based on variation in estimates among pitfall clusters at each site. The model specifies that reference standard conditions are met when the number of amphibian species is equal to or greater than 20. The line shown in the plot is the expected trend based on Figure 7-7, *not* a regression line through the data points.

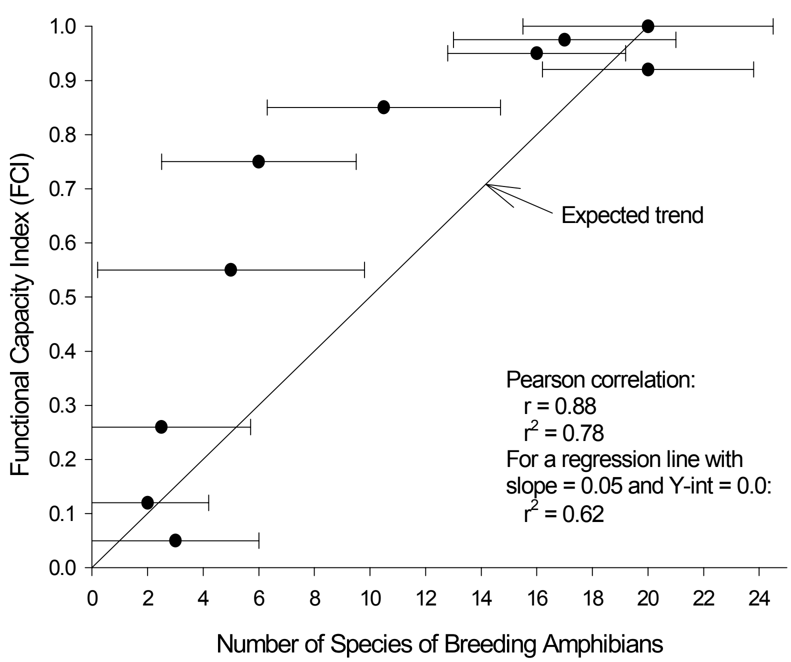

Figure 7-8. Relationship between the FCI determined by applying the model and an independent measure of function (i.e., number of species of breeding amphibians captured in each wetland)

To determine whether the draft model is an accurate predictor of amphibian species richness, the following factors must be considered: (a) the coefficient of determination between FCI and the independent measure of function and (b) the distribution of plotted points in relation to the expected trend. In this example, the coefficient of determination is high and indicates that about 78 percent of the variation in FCI can be accounted for by differences in amphibian species richness. In general, coefficients of determination in excess of 50 percent are desirable, as they indicate that the model is able to account for most of the variance in the two sets of measurements. However, a simple Pearson correlation does not take into account the expected slope or intercept of the relationship. One can achieve a very high correlation between FCI and the independent measure of function and still not be close to the expected trend. One way to evaluate fit of the data points with the expected trend is to use a statistical package (e.g., SAS) to calculate the coefficient of determination between the data points and a regression line whose slope and y-intercept are forced to the expected values. In Figure 7-8, the expected trend has a slope of 0.05 (i.e., 1/20) and intercept of 0.0, and the resulting coefficient of determination is 0.62. Therefore, 62 percent of the variance in the data can be explained by the expected trend.

A simpler but less quantitative way to evaluate fit of the data to the expected trend is by visual inspection of the data plot (Figure 7-8). It can be seen that, despite the relatively high correlation coefficient, the data do not fit the expected trend. Rather, the draft model tends to produce FCI scores that are too high, particularly for sites falling in the middle of the range of the independent measure of function. Some modification of the model is needed to bring these values into line.

There are two ways to modify a draft model to improve its performance relative to the independent measure of function: (a) modify the aggregation equation by changing mathematical functions (e.g., arithmetic means versus geometric means), changing weights or exponents, or by dropping or adding variables; and (b) modify the relationships between the measures of one or more variables and their subindices. Both approaches may be needed to achieve a good fit, and both involve some trial-and-error experimentation.

Say, for example, that the data shown in Figure 7-8 are for a four-variable model of the general form $\text{FCI} = [(V_A + V_B)/2 \times (V_C + V_D)/2]^{1/2}$. One way to improve the correspondence between the data points and the expected trend is to drop the exponent on the aggregation equation, which is equivalent to squaring the right side of the equation. Squaring values of the index does not affect the end points appreciably, since $0^2 = 0$ and $1.0^2 = 1$. However, squaring lowers values in the midrange of the index (e.g., $0.5^2 = 0.25$). Therefore, squaring reduces the curvature of the data plot shown in Figure 7-8 and helps to bring FCI values into line. This modification improves the correlation of FCI versus the independent measure of function to $r = 0.94$ and $r^2 = 0.88$. Regression analysis shows that the fit of the data to the expected trend is now $r^2 = 0.87$.

One type of model modification that should always be considered is dropping one or more variables, particularly if the model contains a total of more than four or five variables. Model simplification by dropping unnecessary or unimportant variables has the added benefit of reducing the amount of time and effort required for users to apply the model and to gather data in the field. Ways to identify variables that might be dropped without reducing model performance appreciably are discussed under "Verifying the Model." Another approach is to examine the relationship between the measure of each variable and its subindex by plotting the average measure for a variable at each site against the independent measure of function (Table 7-4, step 6).

If a variable is important to the performance of the model, then a plot of the measure for that variable versus the independent measure of function should resemble the measure/subindex relationship given in the model. Deviations or outliers should be explainable in terms of the influence of the other variables in the model. For example, Figure 7-9 shows a hypothetical relationship between total organic carbon export (i.e., the independent measure of function) measured in a series of 20 low-gradient riverine wetlands and the percent cover of leaves and fine woody debris determined in sample plots within those wetlands (i.e., variable V_{LITTER} of Ainslie et al. 1999). An FCI of 1.0 corresponds to a carbon export rate of approximately 95 kg/ha/year, the lowest value measured at reference standard sites.

Considerable scatter is expected in the data (Figure 7-9) because the plot fails to consider the effects of other variables that influence carbon export. For example, the points labeled "A" in Figure 7-9 are much higher than expected based on the variable/subindex relationship presented in the model. However, these values might have come from sites where larger woody debris contributes more heavily to organic export. Similarly, the points labeled "B" may be from sites that rarely flood, so that accumulated leaf litter does not contribute greatly to carbon export. The full model contains a variable that accounts for coarse woody debris V_{CWD} and another describing flood frequency V_{FREQ}. Therefore, the outlying values of carbon export ("A" and "B" in Figure 7-9) can be explained based on other variables in the model.

Figure 7-9 suggests that the variable V_{LITTER} is indeed an important factor in carbon export, validating the opinion of the A-Team. Except for deviations that can be explained based on the influence of other variables, the increasing trend in the plot is clear. If the graph were to show a random scatter of points, or some trend opposite to the expected one, it would indicate that (a) the variable may be less important than other variables and might be down-weighted or dropped from the model, (b) the variable may be poorly defined or difficult to measure and should be revised or dropped, or (c) the variable may affect carbon export in some way that the A-Team did not anticipate. In the latter case, the relationship between the variable and its subindex could be redrawn to improve model performance.

Another pattern that might be produced when a response variable (e.g., animal abundance) measured at a number of sites is plotted against a habitat variable is a wedge-shaped scatter of points rather than the more linear pattern shown in Figure 7-9. Recent literature dealing with habitat model testing and

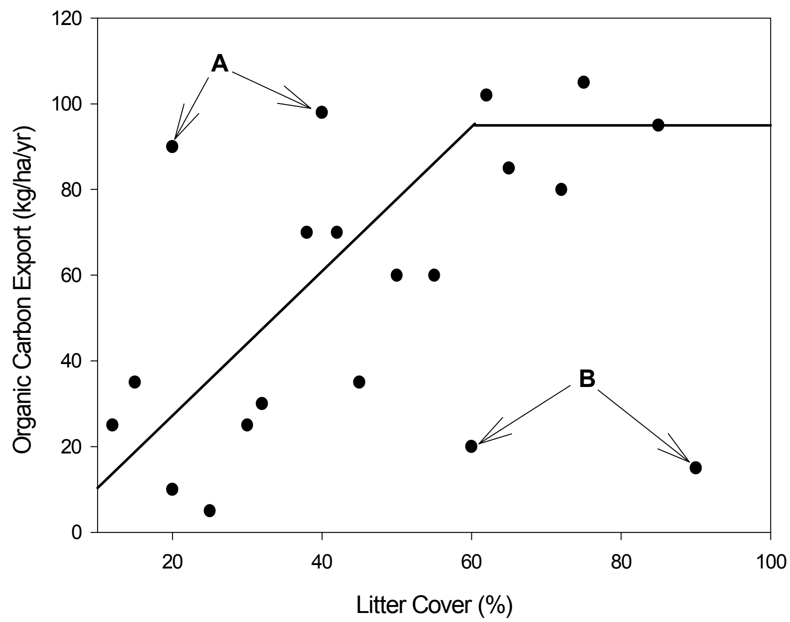

Figure 7-9. Hypothetical relationship between the measure for a variable (i.e., percent cover of leaf litter) and the independent measure of function (i.e., estimates of total carbon export from a number of riverine wetlands). The solid line represents the expected relationship based on the model. See text for details

evaluation (e.g., Terrell et al. 1996, Schroeder and Vangilder 1997) suggests that this wedge-shaped pattern is due to the effects of limiting factors that put an upper limit on the size of a population but may not be important influences on abundance when levels are below this ceiling. Therefore, the important relationship between the response variable and the habitat variable that imposes the ceiling may be represented by the upper surface of the wedge-shaped scatter of points rather than by a regression line through the middle of the cloud. Cade, Terrell, and Schroeder (1999) provide an analytical approach, called regression quantiles, which allows a separate evaluation of species response for populations that are near their maxima (i.e., for data points located near the upper surface of the wedge). This approach may also be useful in evaluating HGM assessment models when relationships between levels of function and environmental variables may be influenced by limiting factors.

After all options for revising the model aggregation equation and variable/subindex relationships have been exhausted, and the effects have been examined by recalculating the overall fit of each new version of the model to the independent measure of function (e.g., Figure 7-8), the result is an altered model that has been forced to fit a particular set of data. Almost certainly, this new model is more accurate and reliable than the draft model. However, the new model also should be subject to validation, using a different set of wetland sites (Table 7-4, Step 8). As stated previously, model validation is an iterative process that continues until the model meets standards for performance demanded by its developers and users.

References

Adamus, P. R. (1995). "Validating a habitat evaluation method for predicting avian richness," *Wildlife Society Bulletin* 23(4), 743-749.

Ainslie, W. B., Smith, R. D., Pruitt, B. A., Roberts, T. H., Sparks, E. J., West, L., Godshalk, G. L., and Miller, M. V. (1999). "A regional guidebook for assessing the functions of low-gradient, riverine wetlands in western Kentucky," Technical Report WRP-DE-17, U.S. Army Engineer Waterways Experiment Station, Vicksburg, MS.

Block, W. M., Morrison, M. L., Verner, J., and Manley, P. N. (1994). "Assessing wildlife-habitat-relationships models: A case study with California woodlands," *Wildlife Society Bulletin* 22, 549-561.

Brinson, M. (1995). "The HGM Approach explained," *National Wetlands Newsletter*, Environmental Law Institute, Washington, DC, 17(6), 7-13.

Cade, B. S., Terrell, J. W., and Schroeder, R. L. (1999). "Estimating effects of limiting factors with regression quantiles," *Ecology* 80, 311-323.

Caswell, H. (1976). "The validation problem." *Systems analysis and simulation in ecology.* B. C. Patten, ed., Academic Press, New York, Vol IV, 313-325.

Cook, J. G., and Irwin, L. L. (1985). "Validation and modification of a habitat suitability model for pronghorns," *Wildlife Society Bulletin* 13, 440-448.

Hall, C. A. S., and Day, J. W., Jr. (1977). "Systems and models: Terms and basic principles." *Ecosystem modeling in theory and practice.* C. A. S. Hall and J. W. Day, Jr., ed., John Wiley and Sons, New York, 5-36.

Kleiss, B. A. (1996). "Sediment retention in a bottomland hardwood wetland in eastern Arkansas," *Wetlands* 16, 321-333.

Lancia, R. A., Miller, S. D., Adams, D. A., and Hazel, D. W. (1982). "Validating habitat quality assessment: An example," *Transactions of the North American Wildlife Conference* 47, 96-110.

Marcot, B. G., Raphael, M. G., and Berry, K. H. (1983). "Monitoring wildlife habitat and validation of wildlife-habitat relationships models." *Transactions of the North American Wildlife Conference* 48, 315-329.

O'Neil, L. J. (1993). "Test and modification of a northern bobwhite habitat suitability index model," Technical Report EL-93-5, U.S. Army Engineer Waterways Experiment Station, Vicksburg, MS.

O'Neil, L. J., Roberts, T. H., Wakeley, J. S., and Teaford, J. W. (1988). "A procedure to modify habitat suitability index models," *Wildlife Society Bulletin* 16, 33-36.

Ott, W. R. (1978). "Structure of environmental indices." *Environmental indices: Theory and practice*. Ann Arbor Science, Ann Arbor, MI, 49-95.

Overton, W. S. (1977). "A strategy of model construction." *Ecosystem modeling in theory and practice.* C. A. S. Hall and J. W. Day, Jr., ed., John Wiley and Sons, New York, 49-73.

Rykiel, E. J. (1996). "Testing ecological models: The meaning of validation," *Ecological Modelling* 90(3), 229-244.

SAS Institute, Inc. (1988). *SAS language guide for personal computers, release 6.03 edition*. Cary, NC.

Schamberger, M. L., and O'Neil, L. J. (1986). "Concepts and constraints of habitat-model testing." *Wildlife 2000: Modeling habitat relationships of terrestrial vertebrates*. J. Verner, M. L. Morrison, and C. J. Ralph, ed., University of Wisconsin Press, Madison, WI, 5-10.

Schroeder, R. L., and Haire, S. L. (1993). "Guidelines for the development of community-level habitat evaluation models," Biological Report 8, U.S. Fish and Wildlife Service, Washington, DC.

Schroeder, R. L., and Vangilder, L. D. (1997). "Tests of wildlife habitat models to evaluate oak-mast production," *Wildlife Society Bulletin* 25, 639-646.

Smith, R. D. (1993). "A conceptual framework for assessing the functions of wetlands," Technical Report WRP-DE-3, U.S. Army Engineer Waterways Experiment Station, Vicksburg, MS.

Smith, R. D., Ammann, A., Bartoldus, C., and Brinson, M. M. (1995). "An approach for assessing wetland functions using hydrogeomorphic classification, reference wetlands, and functional indices," Technical Report WRP-DE-9, U.S. Army Engineer Waterways Experiment Station, Vicksburg, MS.

Terrell, J. W., and Carpenter, J., ed. (1997). "Selected habitat suitability index model evaluations," Information and Technology Report USGS/BRD/ITR-1997-0005, U.S. Geological Survey, Washington, DC.

Terrell, J. W., Cade, B. S., Carpenter, J., and Thompson, J. M. (1996). "Modeling stream fish habitat limitations from wedge-shaped patterns of variation in standing stock," *Transactions of the American Fisheries Society* 125, 104-117.

U.S. Fish and Wildlife Service. (1980). "Habitat evaluation procedures (HEP)," Ecological Services Manual 102, Washington, DC.

__________. (1981). "Standards for the development of habitat suitability index models," Ecological Services Manual 103, Washington, DC.

Waide, J. B., and Webster, J. R. (1976). "Engineering systems analysis: Applicability to ecosystems." *Systems analysis and simulation in ecology.* B. C. Patten, ed., Academic Press, New York, Vol IV, 329-371.

Zar, J. H. (1984). *Biostatistical analysis*. 2nd ed., Prentice-Hall, Englewood Cliffs, NJ.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*
September 2001

2. REPORT TYPE
Final Report

3. DATES COVERED *(From - To)*

4. TITLE AND SUBTITLE
Hydrogeomorphic Approach to Assessing Wetland Functions: Guidelines for Developing Regional Guidebooks
Chapter 7: Verifying, Field Testing, and Validating Assessment Models

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

James S. Wakeley, R. Daniel Smith

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER
32985

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

U.S. Army Engineer Research and Development Center
Environmental Laboratory
3909 Halls Ferry Road
Vicksburg, MS 39180-6199

8. PERFORMING ORGANIZATION REPORT NUMBER

ERDC/EL TR-01-31

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Corps of Engineers
Washington, DC 20314-1000

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Assessment models developed for use under the Hydrogeomorphic (HGM) Approach to assessing wetland functions should be verified, field tested, and validated to ensure that they meet the Assessment Team's goals for accuracy, consistency, and ease of application. Verification is a check of model logic and sensitivity to changes in the input variables. Field testing ensures that users can apply the model efficiently and with consistent results. Validation is the testing of model accuracy by comparing its output against an independent measure of function. This report provides guidance for each of these steps in model testing.

15. SUBJECT TERMS

Application	HGM	Model testing	Validation
Assessment models	Hydrogeomorphic Approach	Sensitivity	Verification

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			**19b. TELEPHONE NUMBER** *(include area code)*
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED		44	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. 239.18