Reconnaissance (1:20,000) Fish and Fish Habitat Inventory: User's Guide to the Fish and Fish Habitat Assessment Tool (FHAT20)

Prepared by

BC Fisheries
Information Services Branch

for the

Resources Inventory Committee

May 2000

Version 1.0

Additional Copies of this publication can be purchased from:
Government Publications Centre
Phone: (250) 387-3309 or
Toll free: 1-800-663-6105
Fax: (250) 387-0388
www.publications.gov.bc.ca

Digital Copies are available on the Internet at:
  http://www.for.gov.bc.ca/ric

Disclaimer

Approval by the Ministry of any deliverable created by this model means only that the deliverables were provided in accordance with standards and specifications of this procedure to the acceptance levels implicit in Ministry quality assurance procedures. Users are cautioned that interpreted information on this product developed for the purposes of the Forest Practices Code Act and Regulations, for example stream classifications, is subject to review by a statutory decision maker for the purposes of determining whether or not to approve an operational plan. (The statutory decision maker is typically the Ministry of Forests District Manager except in areas of joint approval where it is the Ministry of Forests District Manager and the Designated Environment Official).

Abstract

The Reconnaissance Fish and Fish Habitat Inventory is a sampled-based survey covering whole watersheds as defined from 1:20,000 scale maps and air photos. This inventory is intended to provide information regarding fish distribution and relative abundance as well as stream reach and lake biophysical data for interpretation of habitat sensitivity and capability for fish production. While the reconnaissance inventory is intended to cover whole watersheds, time, money, and personnel are not available to survey every stream reach and lake in the watershed; therefore only a subset of reaches and lakes in the watershed is sampled. However, forestry planning processes require the development of products showing the extent of fish distribution or stream channel widths for the entire planning area, not just their distribution in sampled reaches and lakes. These products must be interpreted from the sampled-based inventory. The **F**ish and Fish **H**abitat **A**ssessment **T**ool (FHAT20) is a computer program designed to analyze reconnaissance-level inventory data to produce a set of standardized interpretive products.

FHAT20 is an extrapolation program used to estimate fish habitat characteristics, fish presence and capability in unsampled reaches based on their remote-sensed characteristics and models relating these characteristics to field-based observations in the sampled reaches. FHAT20 uses data stored in the Field Data Information System (FDIS), the standard reconnaissance inventory project database. The end product from FHAT20 is a set of predictions of channel width and probability of fish presence for all reaches. These predictions are used to estimate the most likely Forest Practices Code (FPC) stream class (S1–S6) for each reach and the level of certainty associated with each prediction.

This user's guide documents the background theory and installation and operating instructions for the FHAT20 program. There are seven basic modelling steps that must be followed:

1. Define stratification groups used to make physical predictions;
2. Predict channel width, wetted width, bankfull depth and the probability of non-visible channels for all unsampled reaches, possibly using a stratified analysis;
3. Define fish groupings;

For each fish grouping:

4. Edit feature data to define whether a feature is an obstruction to each fish group;
5. Model the fish group's range within the watershed;
6. Model the fish group's habitat capability which is combined with the predicted range to estimate probability of fish presence; and
7. Model FPC stream classification based on the predicted probability of fish presence and predicted channel widths.

Acknowledgements

This project was funded by contracts from BC Fisheries to Ecometric Research Inc. David Tredger and Tony Cheong, scientific authorities on the project, provided significant input into the design requirements of FHAT20. We thank Geographic Data B.C. for providing digital TRIM maps and Stu Hawthorn at BC Fisheries for providing a digital TRIM watershed atlas for one of the test watersheds. Thanks to Dr. Carl Walters at University of British Columbia for providing the bayesian sampling-importance-resampling algorithm.

Table of Contents

List of Figures

List of Tables

1.0 Introduction

The Reconnaissance Fish and Fish Habitat Inventory is a sampled-based survey covering whole watersheds, i.e., all lakes, stream reaches and connected wetlands within the watershed, as defined from 1:20,000 scale maps and air photos. This inventory is intended to provide information regarding fish distribution and relative abundance as well as stream reach and lake biophysical data for interpretation of habitat sensitivity and capability for fish production (Anon. 1998a). While the reconnaissance inventory is intended to cover whole watersheds, time, money, and personnel are not available to survey every stream reach and lake in the watershed; therefore only a subset of reaches and lakes in the watershed is sampled. However, forestry planning processes require the development of products showing the extent of fish distribution or stream channel widths for the entire planning area, not just their distribution in sampled reaches and lakes. These products must be interpreted from the sampled-based inventory. The **F**ish and Fish **H**abitat **A**ssessment **T**ool (FHAT20) is a computer program designed to analyze reconnaissance-level inventory data to produce a set of standardized interpretive products.

FHAT20 is an extrapolation program used to estimate fish habitat characteristics, fish presence and capability in non-sampled reaches based on their remote-sensed characteristics (derived from 1–20,000 scale maps and air photos) and models relating these characteristics to field-based observations in the sampled reaches. FHAT20 uses data stored in the Field Data Information System (FDIS), the standard reconnaissance inventory project database. The end product from FHAT20 is a set of predictions of channel width and probability of fish presence for all reaches. These predictions are used to estimate the most likely Forest Practices Code (FPC) stream class (S1–S6) for each reach and the level of certainty associated with each prediction.

This user's guide describes how to use FHAT20 and the assumptions and methods of its modelling approaches. Section 2.0 describes how to install FHAT20 and provides an overview of the steps that you must follow to develop standard interpretive products of fish habitat, distribution, and capability. A critical step in using FHAT20 is importing data from FDIS. Section 2.0 describes how this importing procedure works and what to do when errors in the FDIS database are detected.

The remaining sections provide details on the following FHAT20 modules:

1. A physical habitat module which predicts channel width, wetted width, bank full depth, and the probability of non-visible channels in all non-sampled reaches (Section 3.0);

2. A fish grouping module which allows users to summarize species/life-stage specific presence and abundance data collected from the reconnaissance survey into management relevant fish 'groupings' (Section 4.0). These groupings are then used in all subsequent fish distribution and capability modelling. An example of a fish grouping would be all species identified under the Forest Practices Code (FPC) that are used to classify a stream as fish bearing.

3. An obstruction module that allows users to visualize and edit fish obstruction information collected during the reconnaissance survey and from other surveys. These obstructions are used to restrict the range of the fish grouping within the watershed (Section 5.0).

4. A fish range module that allows users to predict a fish group's range in the watershed based on observed fish occurrences, obstruction data, and rules that use predicted or remote-sensed information (Section 6.0).

5. A fish habitat capability module that uses remote-sensed reach characteristics and estimates of relative abundance from the sampled reaches to estimate capability in unsampled reaches (Section 7.0). Fish range and capability predictions for Forest Practices Code fish are combined with channel width predictions to estimate the Forest Practices Code (FPC) stream classification (S1–S6) for each reach.

2.0 Installation and General Operating Guidelines

The FHAT20 computer program consists of: 1) a relational database management system to retrieve data from FDIS and save modelling (extrapolation) results; 2) a series of dialogue boxes and graphics to develop and examine various models predicting physical habitat characteristics, fish range, and habitat capability; and 3) a series of computer algorithms that perform statistical and other extrapolation operations used in the modelling. FHAT20 is a 32-bit application that will operate under Windows95, Windows98, or Windows NT and must be used in conjunction with FDIS version 6.5 or higher.

2.1 Installation

FHAT20 can be downloaded from the INTERNET via the BC Fisheries FTP site. To obtain the installation program go to the FTP site FSHFTP.ENV.GOV.BC.CA. You can do this through an internet browser by using FTP://FSHFTP.ENV.GOV.BC.CA/pub/outgoing/FHAT20 as the address to get into the FHAT directory. Click on FHAT20 ver1.0 May 11 2000.ZIP and you can then download it to your hard disk (e.g. C:\TEMP). Unzip the files into the temporary directory. One of these files is named SETUP.EXE. Run SETUP.EXE to initiate the FHAT20 installation program. There is a file available for download suitable for creating a disk setup version for computers not hooked up to the internet – see the README.TXT file for directions to this and any updates.

When the installation process is complete, you will need to create at least one sub-directory below the directory where you installed the program (e.g., C:\FHAT20\WSHD1). The FDIS database that you want to use to develop an interpretive product using FHAT20 should be copied or moved to this sub-directory. There should be one sub-directory for each FDIS dataset that you wish to analyze. A typical directory structure would be as follows:

DIRECTORY	CONTENTS	SOURCE OF CONTENTS
C:\FHAT20	FHAT20 program	Installation program
C:\FHAT20\WSHD1	FDISDAT.MDB of WSHD1 project	Copy from WSH1 FDIS directory
C:\FHAT20\WSHD2	FDISDAT.MDB of WSHD2 project	Copy from WSHD2 FDIS directory

2.2 First Time Use of FHAT20: Importing a FDIS database into FHAT20

To start FHAT20, double click on FHAT20.EXE from the Windows Explorer. Select the "File/Open Database" menu item, move to the directory that you copied/moved the FDIS database to (e.g., C:\FHAT20\WSHD1) and load FDISDAT.MDB. The first time you run FHAT20 on an FDIS dataset, a set of procedures will be initiated to read the data, check it for errors and omissions, and create a new Access database, MODEL.MDB, which will be used for all future modelling sessions. In subsequent sessions you will still have to move to the appropriate subdirectory and load FDISDAT.MDB, but if you had previously created MODEL.MDB successfully, FHAT20 will actually open MODEL.MDB. MODEL.MDB contains imported FDIS data as well as modelling rules and results that you develop.

FHAT20 uses 7 tables from FDISDAT.MDB to create the MODEL.MDB database: REACH_CARDS; S_SITE_CARDS; FEATURE; FISH_FORM; FISH_GEAR_SPECS; FISH_NET_SPECS; and FISH_EF_SPECS (Fig. 1).

MODEL.MDB contains:
- a subset of information from FDIS used for modelling;
- new variables computed from information in FDIS (e.g., total stream length upstream of each reach);
- a set of default 'rules' used to define fish groupings and model fish range in the watersheds;
- predicted results of channel width, wetted width, bankfull depth, probability of non-visible channels, fish range, and capability computed from previous modelling sessions.
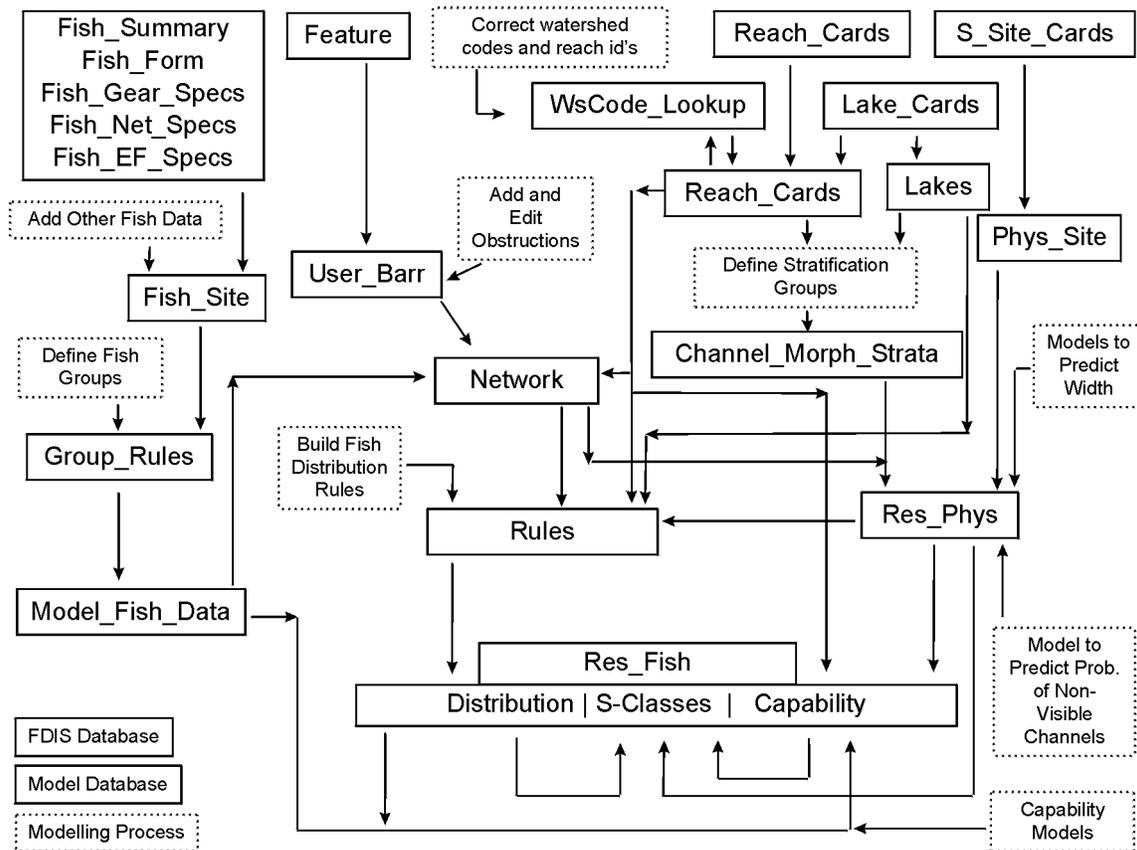
Figure 1: The relationship between FDIS data, various modelling steps performed by the user in FHAT20, and the majority of the FHAT20 model database (MODEL.MDB) where model data, rules, and results are saved.

2.2.1 Error Checking Procedure

When you first load FDISDAT.MDB, FHAT20 automatically checks FDIS data for inconsistencies in the watershed codes and reach ids. These inconsistencies must be corrected because the watershed code and reach id fields are critical to determine the 'connectivity' of reaches and streams within the watershed (the upstream and downstream neighbours of each reach). 'Connectivity' is used to compute a number of variables in the MODEL.MDB database, for example, the total stream length upstream of each reach (used to predict channel width) or whether the reach is upstream of an obstruction or downstream of an observed fish occurrence. The error checking procedure also detects missing values for any FDIS variables that can potentially be used in FHAT20 modelling procedures.

The first step that the FHAT20 import procedure completes is the combination of the Reach_Cards and Lake_Cards tables from FDIS (Fig. 2) into a table called REACH_CARDS in MODEL.MDB. The importing procedure then loops through all reaches in this new table and checks for errors in watershed code and reach_id fields. Errors are reported in the MODEL.MDB table WSCODE_LOOKUP (Table 1). When errors are found, FHAT20 continues with the import procedure, but only includes 'clean' reaches where no errors were detected. Following the import procedure, a dialogue box will appear reporting on the numbers of different types of errors that were encountered. Although it is possible to continue with various modelling steps (described below) to develop interpretive products based on only the 'clean' reaches, it is recommended that the user correct the errors following the procedures outlined below.
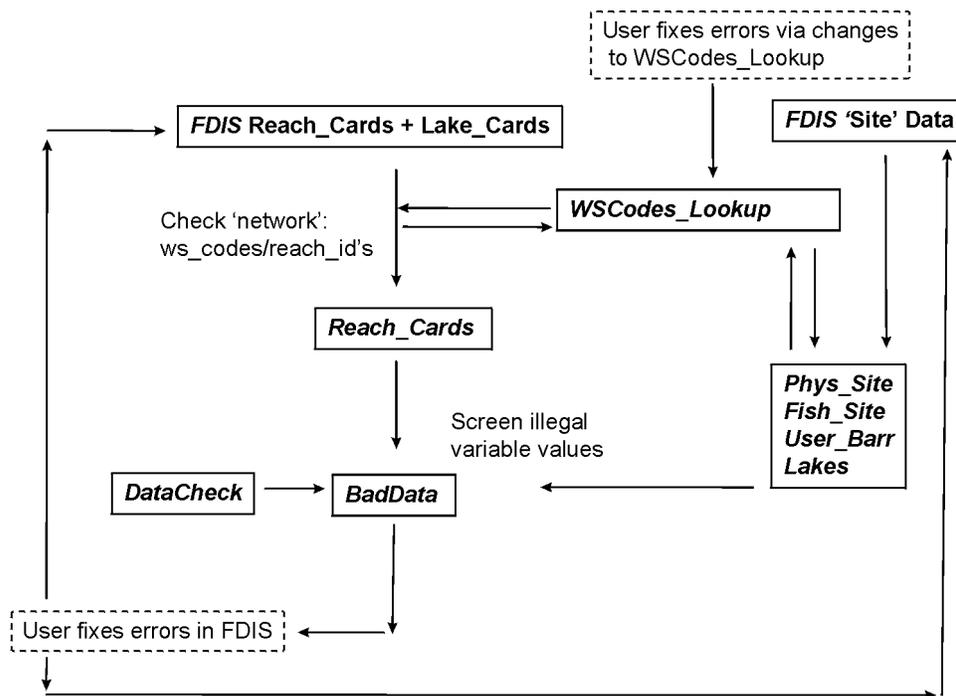
Figure 2: Schematic showing the procedure to import data from FDIS into FHAT20.

Table 1: Description of some of the fields in the WSCODE_LOOKUP table (Fields not listed below are self-explanatory).

Field Name	Field Description
RECID	An original FDIS identifier
FDIS_WS_CODE	The original watershed code from the FDIS database
FDIS_REACH_ID	The original reach_id from the FDIS database
MODEL_WS_CODE	The corrected watershed code used in the MODEL database (you must enter the corrected code in the table)
MODEL_REACH_ID	The corrected reach_id used in the MODEL database (you must enter the corrected id in the table)
ILP	Interim locator point number from FDIS
ILP_MAP	Map sheet number associated with interim locator point
COMMENTS	User defined comments concerning errors and changes
FIXED	True/False denoting that the corrected watershed code and reach id that you entered was accepted during the last import process
IsLake	True/False denoting that the record is a lake (and has a corresponding record in the MODEL Lakes table and the FDIS Lake_Cards table
TABLE_TYPE	If a reach in WSCODE_LOOKUP has corresponding site data (physical, fish, lake, or obstruction), the table name(s) is specified. Users should check that the records in these tables do correspond with the modified watershed code and/or reach_id specified in WSCODE_LOOKUP
ERRORDESCRIPTION	String denoting the type of error and which tables are potentially affected
ERROR_TYPE	An internal code (an integer value, see Table 2) used for error/correction computations in FHAT20 during the importing process. This field only needs to be set by the user to change a watershed code or reach_id of MODEL.MDB site data (to assign it to another reach in the REACH_CARDS record set).

2.2.2 Error Correction

To correct watershed code/reach_id errors, open WSCODE_LOOKUP in Access, fill in the MODEL_WS_CODE and MODEL_REACH_ID fields with the correct values, and save the table. Since WSCODE_LOOKUP will always retain the original FDIS (FDIS_WS_CODE, FDIS_REACH_ID) and corrected (MODEL_WS_CODE, MODEL_REACH_ID) values, you will always be able to link the model results (interpretive products) back to the original FDIS database should the need arise. When the corrections have been made to WSCODE_LOOKUP, close Access and select the "Rebuild Database" choice from the "Utilities" main menu item to re-initiate the import process. Rebuilding the database incorporates all the changes made in WSCODE_LOOKUP into the new MODEL.MDB database. Hopefully all errors will be corrected during the MODEL.MDB rebuilding process. If errors are still reported, repeat the sequence just described, but this time only correcting records in WSCODE_LOOKUP where the FIXED field = FALSE. Records where the FIXED field = TRUE were corrected during previous rebuilding events. This cycle can be repeated as many times as required. The error reporting form displayed at the end of each import process always provides a summary of the current state of the data in terms of how many errors were originally detected and how many have been corrected.

You can view the error report at any time by selecting the "Error Reporting" option from the "Utilities" main menu item (Fig. 3).

Figure 3: The error reporting dialogue box displays a record of errors and corrections made to the FHAT20 model database.

Correcting errors in watershed codes and reach ids for each record in the WSCODE_LOOKUP table in MODEL.MDB requires an understanding of the types of errors that can be trapped and handled by the importing procedure. These errors are distinguished based on text in the COMMENTS field of WSCODE_LOOKUP (Table 2).

It is important to note that in MODEL.MDB, the watershed code and reach id fields link records in REACH_CARDS with records in the other tables containing site data (PHYS_SITE, FISH_SITE, USER_BARR, LAKES). Making a change to a watershed code or reach id in REACH_CARDS (via modification of WSCODE_LOOKUP) may result in two situations:

a) A record in a site' table may no longer have a corresponding record in REACH_CARDS (in which case error Types 4–7 will be reported in WSCODE_LOOKUP) or;
b) A record in a site table may match a record in the corrected REACH_CARDS table, but the site data does not belong to that REACH_CARDS record and instead is actually associated with another reach.

Table 2: Description of error codes reported in WSCODE_LOOKUP.

Code	Error Description
0	*Invalid Watershed Code:* Watershed code is invalid (e.g., 000-...., 999-, NA)
1	*Duplicate Reach ID*: Two records in the FDIS table Reach_Cards have the same watershed code and reach_id. This likely resulted from: a duplication resulting from the importing of watershed codes into FDIS[a]; or an incorrect specification of reach_id or watershed code when data were entered in FDIS.
1	*Duplicate Reach ID (reach is a Lake)*: A record in the FDIS Reach_Cards table matches (same watershed code and reach_id) a record in the FDIS Lake_Cards table. In this case, the "isLake" field in WSCODE_LOOKUP will be "TRUE." This error type is distinguished from the previous error type only because it may be helpful to know which record is the lake when attempting to determine the correct reach_id (or watershed code) for one of the records.
2	*No Parent stream from FDIS ('orphan')*: With the exception of the mainstem, every stream (watershed code) in the FDIS database must have a stream that it flows into (hereafter referred to as its 'parent' stream). This error code denotes that a record (a unique watershed code and reach id) does not have a parent stream and that it is not the mainstem; it is therefore an 'orphan' reach.
3	*Stream eventually flows into a stream with no parent ('incomplete lineage').* The current stream flows into a parent stream (it is not an 'orphan' or the mainstem), but there is a break in the network (a "non-parent" situation) somewhere between its parent and mainstem.
4	*Record in PHYS_SITE with no matching record in Reach_Cards:* The PHYS_SITE table in MODEL.MDB is populated based on records in the FDIS table S_SITE_CARDS. When you correct a watershed code or reach id in REACH_CARDS (MODEL version) by making changes in WSCODE_LOOKUP and re-running the import procedure, you can create a situation where some records in PHYS_SITE have no match in REACH_CARDS in the MODEL database. This error can also arise from incorrect entry of watershed codes or reach ids when entering site data in FDIS.
5	*Record in FISH_SITE with no matching record in Reach_Cards:* The FISH_SITE table in MODEL.MDB is populated based on watershed codes and reach ids in the FDIS table Fish_Form. When you correct a watershed code or reach id in REACH_CARDS (in MODEL version) by making changes in WSCODE_LOOKUP and re-running the import procedure, you can create a situation where some records in FISH_SITE have no match in REACH_CARDS in the MODEL database. This error can also arise from incorrect entry of watershed codes or reach ids when entering fish data in FDIS.
6	*Record in LAKES with no matching record in Reach_Cards:* The LAKES table in MODEL.MDB is initially populated based on watershed codes and reach ids in the FDIS table Lake_Cards and these records are merged into the MODEL.MDB Reach_Cards table. When you correct a watershed code or reach id in REACH_CARDS (in MODEL version) by making changes in WSCODE_LOOKUP and re-running the import procedure, you can create a situation where some records in LAKES now have no match in REACH_CARDS in the MODEL database.
7	*Record in USER_BARR with no matching record in Reach_Cards:* The USER_BARR table in MODEL.MDB is initially populated based on watershed codes and reach ids in the FDIS table FEATURE (linked to S_Site_Cards). When you correct a watershed code or reach id in REACH_CARDS (in MODEL version) by making changes in WSCODE_LOOKUP and re-running the import procedure, you can create a situation where some records in USER_BARR have no match in REACH_CARDS in the MODEL database. This error can also arise from incorrect entry of watershed codes or reach ids when entering obstruction data in FDIS.

[a] Errors in watershed codes are generally associated with the process of importing the watershed codes in the ILP table back into FDIS. In some instances, in headwater forks, when the mainstem and tributary gets assigned a code which is opposite to what was identified, this results in both a missing watershed code and a duplicate watershed code and reach number. Newer versions of FDIS will check for this problem.

Correction of situation b) errors cannot be done automatically. If a match between a corrected record in REACH_CARDS and a record in a site table occurs, the import procedure has no way of knowing whether the identifiers were originally entered correctly in the site table via FDIS (and should therefore now be linked to the record in REACH_CARDS with the original identifiers) or whether the site identifiers were entered incorrectly from FDIS (and should therefore be updated based on the corrected values specified in WSCODE_LOOKUP). Thus, when situation b) occurs during the import procedure, the TABLE_TYPE field in WSCODE_LOOKUP is set to the name(s) of the site table(s) where a record corresponding to the modified REACH_CARDS record exists. It is up to the user to then manually check these cases (using original maps, UTMs, etc.) to make sure that the records do correspond. If they do not, additional records must be added to WSCODE_LOOKUP by the user so that the records in the MODEL.MDB site tables are updated. These new records should contain the original watershed code, reach id, and RecID identifiers (the FDIS_ values, just copy their values) and the new identifiers (MODEL_ values) that are used to link it to the correct REACH_CARDS record. In addition, the ERROR_TYPE field should be set to the appropriate code (e.g., 4 if you are updating a record in the PHYS_SITE table in MODEL.MDB).

Correction of stream network errors detected by FHAT20 may require a significant amount of effort and depends completely on the consistency in the FDIS database. The correction procedure can be streamlined by realizing the hierarchical nature of some of the errors. You may observe a large number of records with 'incomplete lineage' errors (type 3) all resulting due to a small number of 'orphans' (error type 2). If you correct type 'orphan' errors first, all 'incomplete lineage' errors will automatically be corrected by the import procedure. The most efficient way to correct the data is to:

- Correct all 'orphan' type errors (type 2) and known problems with watershed codes (type 0) first;
- then correct duplicate reach id errors (type 1) which may involve changing watershed codes (i.e., the reach id may be correct and the error resulted from an error in the watershed code);
- correct error types 4–7; and finally,
- check and make sure that records in PHYS_SITE, FISH_SITE, USER_BARR, and LAKES are correctly assigned to the new watershed codes or reach ids in REACH_CARDS (check all records with values in the TABLE_TYPE field in WSCODE_LOOKUP). If there are errors, follow the procedure specified in the previous paragraph (add new records to WSCODE_LOOKUP).

Never correct 'incomplete lineage' errors, they can only be eliminated by correcting 'orphan' errors and are only provided in WSCODE_LOOKUP in case you run the modelling procedures without completing all corrections (thereby providing a catalogue of reaches missing from the analysis).

Note it is possible to run the modelling procedures with errors in the watershed codes and reach ids, however none of these reaches will be included in the model analysis. At the very least, you have a record of these reaches in the WSCODE_LOOKUP table (all records with the FIXED field = "FALSE"). However, predictions for other reaches still included in the analysis may be effected by the deletion of the problem reaches due to network connectivity aspects of the modelling. **Thus, running the model with missing reaches is *not recommended*,** especially if they are 'parents' and have reaches flowing into them (i.e., not first order, headwater reaches).

Errors in the watershed code that do not result in 'orphan' streams are not detected during the FHAT20 import procedure. There are two types of such errors that could affect modelling results:

Watershed Code Does Not Reflect Position of Confluence Along the Parent Stream

A tributary of a parent stream may have a code specifying that its confluence is mid-way up the parent (e.g., 120-907500-63000-50000, located 50% of the way up its parent stream 120-907500-63000). However, during inspection of the FHAT20 electronic map (plotting variables determining which tributaries are not accessible to fish because of obstructions) you may notice that the confluence of the child stream is actually further upstream that specified by the watershed code. Assuming the map is correct, such an error could result in erroneous predictions. For example, an obstruction in the parent stream located 60% of the way up its total length (on the map) would not limit access to the tributary, because electronically, the obstruction is upstream of the tributary confluence, which is only 50% of the way up the parents total length. To correct this type of error, you need to estimate the correct watershed code for the tributary in question based on the actual proportional distance of its confluence relative to the parents total length (e.g., 75%). A new record in WSCODE_LOOKUP must be added containing the original FDIS and modified (MODEL) watershed codes (e.g., 120-907500-63000-50000 and 120-907500-63000-75000, respectively).

Watershed Code Does Not Reflect Stream Hierarchy

A stream may appear electronically to be a tributary of a particular parent stream based on its watershed code, when in reality it is not a tributary of that parent stream (or vice-versa). For example, 120-907500-63000-50000-0123 would electronically be treated as a tributary of 120-907500-63000-50000. However, on inspection of the map, the stream in question may actually be a tributary of 120-907500-63000. Its watershed code needs to be modified in WSCODE_LOOKUP to reflect its actual position in the stream hierarchy and its proportional confluence distance relative to its parent stream (120-907500-63000-57000).

During the importing routine, FHAT20 catalogues records with missing values for variables that can be used in the modelling process. There will be a unique record in the BadData table of MODEL.MDB for each field with a missing value for a given reach (i.e., there can be multiple records per reach). If you want to enter valid values for the records with missing values identified in BadData, you must enter them in FDIS and then re-import the data using the "Rebuild Database" option from the "Utilities" main menu item (Fig. 2). Note that it is possible to continue modelling with missing values, however FHAT20 cannot make predictions for reaches with missing values for a variable if that variable was included in one of the model(s). The table DataCheck in MODEL.MDB contains the list of variables in FDIS that will be screened by the 'data checking' procedure during the import process. You can add variables to this list, or deactivate variables (by unchecking the IsActive field) so they are no longer screened for missing values. DataCheck allows the user to modify the range of legal values that a field can have, and what to do if a particular record does not meet these criteria. Records that fail the data screening procedure for a given field are either deleted, or the missing value is replaced with a missing value flag (-9999 for numeric fields or 'NS' for text fields). The is Dropped field in the BadData table specifies which of these two actions will be taken.

2.2.3   FHAT20 Maps: Stick Diagrams or the Digital TRIM Atlas

A key component of FHAT20 is the ability to view some FDIS data and all model predictions on a digital map of the study area. Two types of maps can be used by FHAT20: 1) a 'stick diagram' where each reach is a straight line between the upstream and downstream reach coordinates; and 2) a TRIM atlas which is a digital version of the hardcopy TRIM maps with watershed codes and reach measures (the distance of each reach break from the confluence of each stream) for each reach in the FDISDAT database.

Stick Map

The 'stick' map represents each reach in the study area as a single, straight line. Such diagrams do not contain bends, lake boundaries, and other features that will be shown on the TRIM atlas maps. Stream reaches are shown as straight lines while reaches classified as lakes (i.e., with records in the FDIS Lake_Cards table) are shown as fatter lines or filled circles (if the lake length could not be computed). FHAT20 uses the UTM coordinates from the REACH_CARDS table to create this 'stick' diagram during the import process and saves this information to the file MODEL.MIF in the project sub-directory. The stick map is actually a MAPINFO file that can be viewed in FHAT20 or MAPINFO. When you view this map for the first time in FHAT20 after the import procedure, you will note that the first reach of each stream 'hangs' in the sense that it does not connect with its parent stream. Such reaches will be represented by open circles at their upstream boundaries. This occurs because FDISDAT.MDB does not store the UTM coordinates for the downstream end for the first reach of each stream. Some of these coordinates can be obtained from the 1:50,000 Watershed Atlas (although coordinates may differ significantly from those on a 1:20,000 map) and others may have been created by the ILP/watershed code process. If you have some or all of these 'first node' coordinates, create a comma-delimited file with 3 fields: watershed code (45 digits – can contain hyphens or not) and UTM easting and northing coordinates in the following format:

```
WS_CODE,EASTING,NORTHING
120907500000000000000000000000000000000000000,674071,5621503
120907500045320000000000000000000000000000000,674810,5622958
120907500045325767200000000000000000000000000,674344,5623477
```

From the "Utilities" main menu item, select the "Import First Nodes File" option. This will read in the .CSV file you created and add these records to the FIRSTNODE table in MODEL.MDB. This process will add the new coordinates to the 'stick' map eliminating some of the 'hanging' tributaries and making the map more readable.

TRIM Watershed Atlas Map

If a digital TRIM Watershed Atlas is available for the study area (with watershed codes that correspond to those in FDISDAT.MDB), a method exists to bring TRIM linework into the model. However, as the TRIM Watershed Atlas is not available provincially, and this procedure is under development, it has not been included in this manual. The 'Load Map File' and 'Load Lakes File' under the 'Edit Legend' selection in "Utilities" are used in this process.[1]

---

[1] For information on the status of the TRIM Watershed Atlas and the procedure to import TRIM linework into the model, please email the BC Fisheries Information Services Branch at RECCE@gems2.gov.bc.ca.

2.3 Controlling the Map Display

The main FHAT20 form (window) consists of a digital map of the watershed that is used to display reach specific data or model results using a color-code system (Fig. 4). Also shown is a legend depicting what the color codes represent, and a set of combo boxes that are used to select various data/model results to display. Data/results are organized into four functional variable groupings shown in the upper combo box. Within each grouping there are a series of variables which are shown in the combo box labeled "Variables" (each variable generally corresponds to a unique field in tables within MODEL.MDB). Table 3 provides a listing of the functional data groupings and their fields, and what each field represents. To view results for any field, simply select the appropriate grouping and field from the combo boxes.
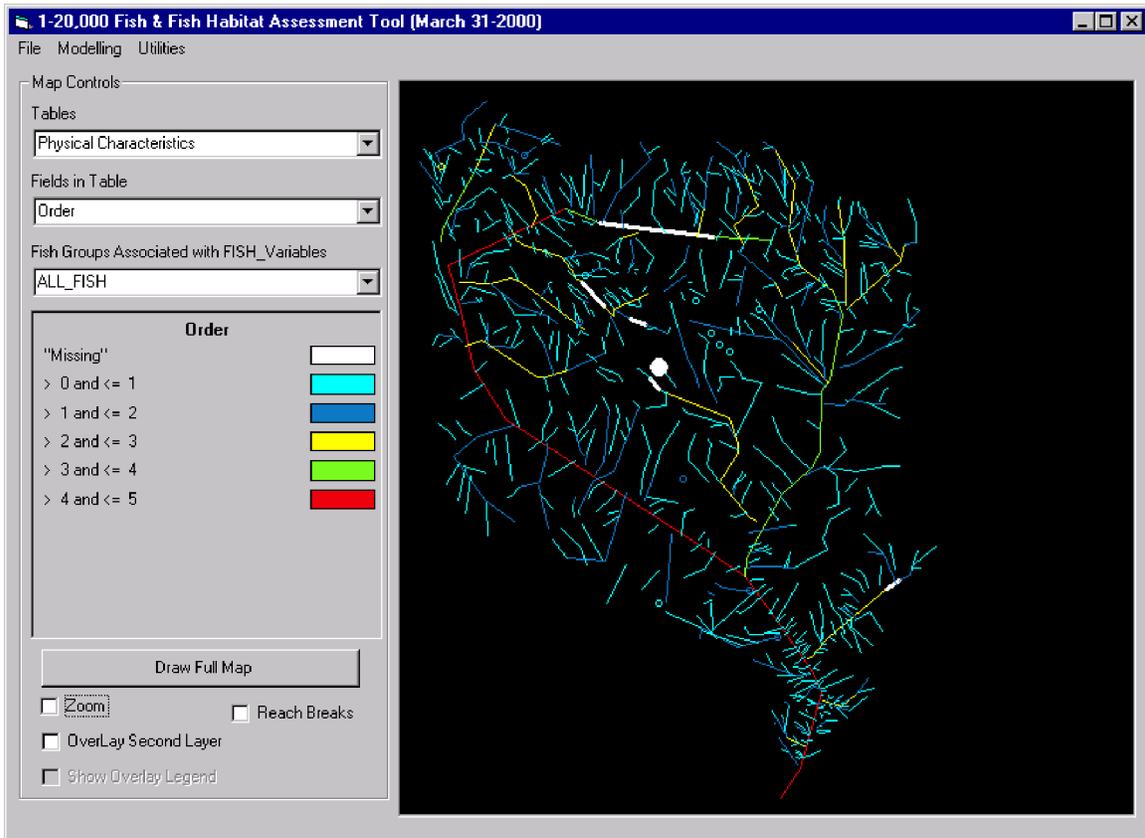
Figure 4: The main FHAT20 dialogue box used to display and query reach-specific data and model results.

The mapping display has a number of features to facilitate data review and analysis.

Point-and-Click to get Reach Attribute Information

Double-click on any reach with the left mouse button to bring up the "Search Results" dialogue box (Fig. 5). A table shows the watershed code and reach id for the reach you selected and the value for the current layer on the map. You can then find the value for other reach attributes by selecting the variable group from the combo list box in the dialogue box and double-clicking on the variable. Click on the watershed code in the table to highlight the stream reach on the map (the selected reach will be displayed as a thick white line).
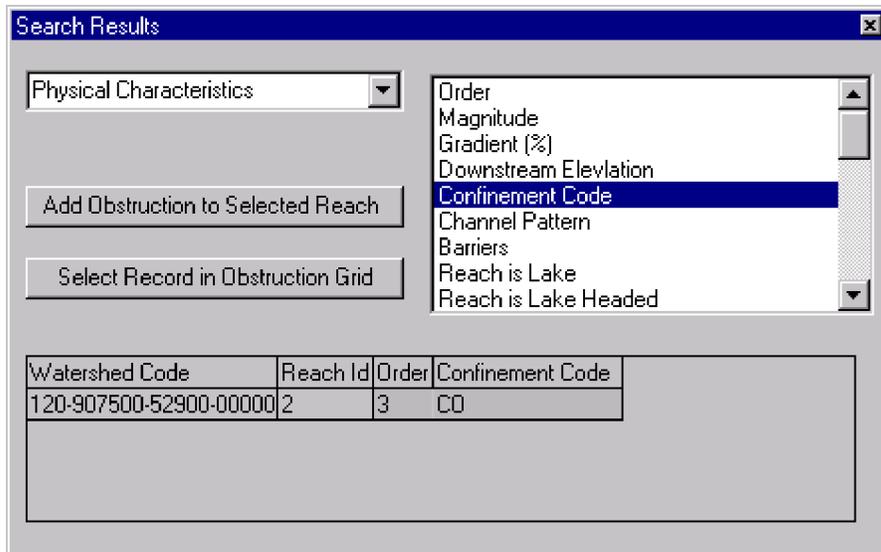
Figure 5: The search results dialogue box displays data and results for reaches selected on the map.

'Zooming-In' on a Selected Area on the Map

The zoom feature allows you to view a smaller area within the full study watershed shown on the map. Simply click on the check box labeled 'zoom', click on the map with the left mouse button, and hold the button down while dragging the mouse over the portion of the map that you want to select. When you release the mouse button the map will be redrawn showing only the selected area. You can repeat this process to select an even smaller area to view, or restore the map to its full size by clicking on the button labeled "Draw Full Map." You can view the reach breaks by clicking on the check box labeled "Show Reach Breaks."

Overlay Second Layer - Looking at Two Map Variables at One Time

The overlay feature is very useful for comparing predicted results with observed sample data. This feature allows you to look at two different variables on the map at one time. First select a variable and display it on the map, click on the check box labeled "Overlay Second Layer," and then select the second variable you want to display. The second variable will be displayed as color-coded squares shown in the center of each reach on top of the first variable. Click on the "Show Overlay Legend" check box to bring up legends for both layers simultaneously.

Table 3: Organisation of variables (data and model results) that can be displayed in the FHAT20 Map Interface. The Group and Variable Name columns correspond to the items shown in the combo boxes labelled "Group" and "Variable" on the FHAT20 main window. The Table/Field column gives the table and field name where the data/results are stored in MODEL.MDB. 'Group' in the table is used to designate a fish grouping name (e.g., FPC_FISH); there will be multiple occurrences of fields preceded with 'Group', one set of each fish group modelled.

Group Name	Variable Name	Description	Table/Field
Physical Predictions	Order	Strahler stream order	Reach_Cards / Order_20
	Magnitude	Magnitude of reach (# of 1st order streams upstream of reach)	Reach_Cards / Magnitude_20
	Gradient	Reach gradient	Reach_Cards / Gradient_20
	Downstream Elevation	Elevation (metres above MSL) of downstream node of reach	Reach_Cards / Downstream_Elevation
	Confinement Code	Confinement of reach	Reach_Cards / Conf_code
	Channel Pattern	Channel pattern of reach	Reach_Cards / Cptn_Code
	Features	Location of Features (from FDIS Features table or user-defined)	User_Barr
	Reach is a Wetland	Denotes whether reach is a wetland	Reach_Cards / Wetland
	Reach is U/S of a Wetland	Denotes which reaches are upstream of wetlands	Reach_Cards / USofWetland
	Reach is a Lake	Denotes whether the reach has a corresponding record in the Lakes table	Reach_Cards / IsLake
	Reach is Lake-headed	Denotes whether a reach has a lake upstream	Network / IsLakeHeaded
	Reach flows into Lake	Denotes whether a reach is an inflow stream to a lake	Network / ReachFlowsIntoLake
	Upstream Length	Length (km) of stream upstream from the downstream end of each reach	Network / Uplen
	Maximum Downstream Gradient	The maximum gradient between the reach and the first reach at the most downstream end of the watershed	Network / Maxdsgrade
	Parent Order	The Strahler order of the parent stream of the first reach of its tributary	Network / Parent_order
	Sampled for Physical/Fish Data	Was the reach sampled for physical parameters or fish (True/False)	Network / Sampled_Phys Network / Sampled_Fish

Table 3. Con't.

Group Name	Variable Name	Description	Table/Field
Physical Predictions	Predicted Channel Width	Predicted Channel Width (m) if the reach is not sampled, Average measured channel width if the reach has been sampled	Res_Phys / Chan_width
	Distance to Bottom of System (km)	Distance of reach (km) to the most downstream reach in the project area	Network / Dist_To_Bottom
	Sinuosity	Ratio of reach length to the straight line distance between the bottom and top of the reach	Reach_Cards / Sinuosity
	Predicted Wetted Width	Same as above but for wetted width	Res_Phys / Wet_width
	Predicted Bankfull Depth	Same as above but for bankfull depth	Res_Phys / Bank_Full_Depth
	Probability of No Visible Channel	The predicted probability that the reach is classified as a non-visible channel	Res_Phys / Novis_prob
	Observed Non-Visible Channel	Denotes whether a sampled reach was a non-visible channel	Phys_Site / No_Channel_Vis
	'Group' Upstream of Obstructions	Reaches upstream of obstructions for the fish group (False-fish group is not upstream of an obstruction specific to that group)	Network / 'Group'_Dsbarr'
	'Group' Downstream of Occurrence	All reaches downstream of an observed occurrence of the fish group	Network / 'Group'_Ds_Fish_Pres
	'Group' Dist U/S of Fish Occurrence	Distance upstream (km) from nearest fish occurrence	Network / 'Group'_Dist_US_Fish_Pres
	'Group' Not Present Upstream	All reaches upstream of sampled reach where fish group was not found (as long as there is not a fish occurrence for group upstream of this point)	Network / 'Group'_US_Fish_Abs
	'Group' Parent D/S of Fish Occurrence	Reach is a tributary of a reach that is downstream of a known fish occurrence	Network / 'Group'_Parent_DS_Fish_Pres
Observed Fish Data	'Group' Reach Presence/Abundance	Denotes fish presence if the site was not electrofished, and density (#/100 m^2) if the site was electrofished	Based on a combination of variables in Model_Fish_Data / (Meth, TotalNo & Cpue)

Table 3. Con't.

Group Name	Variable Name	Description	Table/Field
Fish Predictions	'Group' Range	The maximum range of the fish group in the watershed	Res_Fish / 'Group'_Dist
	'Group' Most Probable Capability Class	The most probable fish habitat capability class	Res_Fish / 'Group'_Most_Prob_Class
	'Group' Probability of No Capability	The probability that the reach has no capability (that the abundance is less than 1 fish in the sample site area)	Res_Fish / 'Group'_Prob_Absent
	'Group' Probability of Low Capability	The probability that the reach has low capability	Res_Fish / 'Group'_Prob_Low
	'Group' Probability of Medium Capability	The probability that the reach has medium capability	Res_Fish / 'Group'_Prob_Medium
	'Group' Probability of High Capability	The probability that the reach has high capability	Res_Fish / 'Group'_Prob_High
	'Group' Probability of Presence	The probability that reach is utilised by the fish group	Res_Fish / 'Group'_Prob_Pres

Table 3. Con't.

Group Name	Variable Name	Description	Table/Field
Stream Classification	Most Probable Stream Class	The most probable stream classification based on the combined probability of different stream widths and discontinuous fish presence (a reach may have a lower probability of fish presence compared to an upstream reach)	Res_Fish / Most_Prob_S
	Uncertainty in S Class	A relative index of uncertainty associated with the Most Probable Stream Class. A value of '0' denotes the minimum uncertainty; 100% probability for one of the S Classes. A value of '100' denotes maximum uncertainty; all S Classes have an equal probability (16.67% because there are 6 classes).	Res_Fish / S_Uncert
	Sx Probability (x=1 – 6)	The probability that the reach is an Sx stream class (e.g., S1) based on combined probabilities	Res_Fish / Sx
	FPC Stream Class	The stream class based on the most likely width, with fish presence determined from FPC_Fish_Present	Res_Fish / ConSClass
	FPC Fish Present	Presence/absence determined from continuous probability of presence values combined with a minimum probability of presence limit defined by the user	Res_Fish / ConFishPres
	Continuous Probability of Presence	Probability of fish presence defined on a continuous basis (downstream reaches cannot have probabilities less than upstream reaches)	Res_Fish / ConPop

Controlling the Map Legend

Each variable that can be mapped has a unique legend that controls how the data will be displayed on the map. If you double-click the mouse on the colored boxes in the legend, the "Legend" dialogue box will be displayed (Fig. 6). Alternatively, select the "Edit Map Legend" choice from the "Utilities" main menu item. There are a number of parameters that you can change to adjust the legend. To display a variable via color codes on a map, a set of bins must be established for specific ranges of the variable (e.g., bin 1 = gradient 0–3%, bin 2 = gradient 3%–7%, bin 3 = …). There is a corresponding color for each bin.

- You can alter the number of bins (labeled number of strata in the dialogue box) and manually set the lower and upper end of the range for the variable.
- Alternatively, you can click on the button labelled "Set Upper and Lower Ranges" to have the program determine the minimum and maximum value for the variable.
- After you have made any of these edits click on the button labelled "Ramp Breaks" to ramp the breakpoints between the lower and upper range values.
- To change the colour associated with any bin, click on a colour in the colour palette and then click on the coloured box adjacent to the bin.
- You can manually edit the bin breakpoints by editing the values in the breakpoint text boxes.

If you wish to edit the legend for another variable simply select it from the list. If you are editing the legend for a variable currently displayed on the map, you must click on the button labelled "Apply Legend" to see the changes. Note that any changes you make to the legend are saved to a file called MODEL.MDE and effect the display in all subsequent sessions.
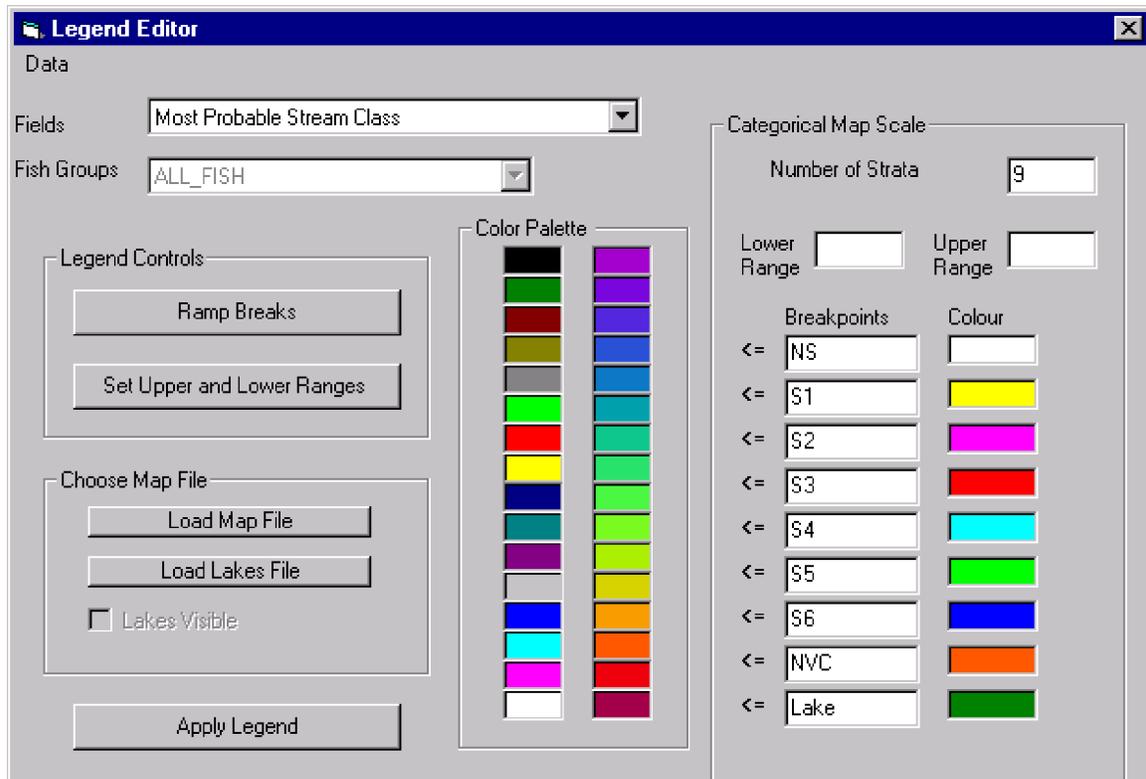
Figure 6: The legend editor dialogue box is used to edit the map display of data and predicted variables.

2.4 General Modelling Procedure

A series of specific steps must be followed to develop standard interpretative products of fish habitat, distribution, and capability using FHAT20. As you develop models and apply them, modelling rules and results are saved to various tables in the FHAT20 database (MODEL.MDB). The relationship between FDIS data, various modelling steps, and the FHAT20 database is shown in Figure 1. There are seven basic modelling steps that must be performed in a specific order (Fig. 7):

1. Define stratification groups used to make physical predictions (Section 3.2);
2. Predict channel width, wetted width, bankfull depth (optional) and the probability of non-visible channels for all unsampled reaches, possibly using a stratified analysis (Section 3.0);
3. Define fish groupings (Section 4.0);

For each fish grouping:

4. Edit obstruction data to define whether a feature is an obstruction to each fish group (Section 5.0);
5. Model the fish group's range within the watershed (Section 6.0);
6. Model the fish group's habitat capability (Section 7.0) which is combined with the predicted range to estimate probability of fish presence; and
7. Model FPC stream classification (Section 8.0) based on the predicted probability of fish presence and predicted channel widths.

A change in data or modelling assumptions in any step requires that all steps following that point are reprocessed. For example:
- if you modify the relationships predicting channel width, and have fish range rules that depend on channel width, you will need to rerun these rules (i.e., repeat step 5).
- If you modify channel width or wetted with predictions you must rerun the habitat capability calculations.
- If you modify channel width you must rerun the habitat capability and stream classification procedures.
- If you add or edit an obstruction that affects the range of a fish group, you will need to rerun its range rule, and since the habitat capability calculations are dependent on the fish range, you will also have to rerun the capability calculations.
- If you rerun habitat capability for a fish group, you will need to rerun the stream classification if it was run using the same fish group. The habitat capability modelling contributes to the calculation of the probability of fish presence variable, which is used in the stream classification procedure.
- If you define a new fish grouping, you must complete steps 4–6 (and possibly 7) for this group.
- If you modify the original data in FDIS, you must re-import the data and reprocess all the results.
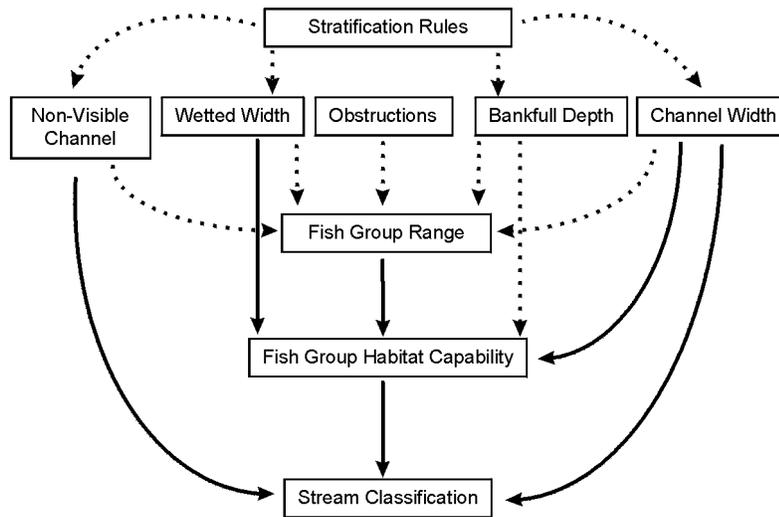
Figure 7: Schematic showing the relationship among modelling steps in FHAT20. Solid lines denote fixed relationships (e.g., channel width is used to predict stream class), while dashed lines show relationships that depend on whether particular variables are used in later modelling steps (channel width depth may be used in predicting fish range, but it is not mandatory).

To assist you in following the correct order in the modelling procedures, FHAT20 keeps track of the sequence and time stamps of all modelling operations performed. A dialogue box, accessed from the "Show Operational Tracking" choice below the "Utilities" main menu item displays this information (Fig. 8) The dialogue box consists of a grid with rows for each modelling step and columns for each of the currently defined fish group. When the data is first imported, all the cells in the grid will be purple, denoting that none of the operations have been performed. As you begin to perform various modelling tasks, the appropriate cells on the grid turn green and the date/time stamp that the task was performed is shown. If you redo a particular model step (e.g., predict channel width), all modelling procedures that depend on that operation will be out of date. These steps will need to be redone, and are depicted as red cells on the grid.

Figure 8: The operational tracking dialogue box displays the status of modelling procedures and the date/time the procedures were run.

You can easily navigate from one modelling procedure to another by double clicking on individual cells in the grid. If you attempt to perform a task out of sequence, that is, perform a task that depends on a previous model step that has not been run (purple cell) or is out of date (red cell), FHAT20 will warn you and stop you from performing the operation. If you want to disable this, toggle the "Operation Tracking" check box off (or deselect the "Operational Tracking" choice below the "Utilities" main menu item).

Due to the structure of the grid, some operations appear fish group specific but are not. Note that there are physical prediction rows (channel width, wetted with, probability of non-visible channel, bank full depth) for each fish group depicted on the grid, but these events are not fish group specific. When you compute one of these variables you will see that the appropriate cell in the grid is changed for all fish groups. If you edited a model result (section 9.0) and later recalculate that result, the "Results Edited" cell will show that the edit rule is out of date. The edit rule itself is not out of date, however this tells you that you had previously edited the results from a particular operation, have since recomputed the operation, but have not re-edited the result. The grid alerts you to this fact, but does not block you from performing additional operations that depend on the result. This is logical as you may not need to edit the new result.

2.5 Exporting Results, Saving Maps, Printing Maps

Model results can be saved to the EXPORT table in MODEL.MDB for analysis in other applications and display in FDIS Map. To export results, select the "Export Results" option from the "File" main menu item. The EXPORT table contains a number of identifier fields to facilitate linkage back to FDIS and other BC Fisheries Inventory applications including:

- FDIS watershed code and reach_id
- MODEL.MDB watershed code and reach_id (equivalent to FDIS values unless the values were corrected in WSCODE_LOOKUP)
- NID, NID_MAP

- ILP, ILP_MAP
- Easting and Northing UTM

When you export the model results, a table called CUREXPORTLOG is updated. This table contains the rules (SQL statements), results, and time stamps associated with the various modelling operations you completed. **This allows a third party to verify that model results in the EXPORT table were based on modelling steps completed in the correct sequence.**

Maps displayed in FHAT20 can be exported as bitmap files. Select the "Dump Map/Legend to Bitmap" option form the "File" main menu item. You will be prompted for a filename to save the map to and a separate filename to save the legend to. You can import both of these files into another application, and because the map and legend images are saved in different files, you have flexibility in terms of where the legend is located on the final graphic. You can also print the map and legend to the default printer by selecting the "Print Map/Legend" option from the "File" main menu item.

3.0 Physical Predictions

FHAT20 predicts channel width, wetted width, and bankfull depth for all unsampled reaches and the probability that these reaches are non-visible channels.

3.1 Channel Width, Wetted Width, and Bankfull Depth Predictions

FHAT20 predicts channel and wetted width and bankfull depth in unsampled reaches as a function of their upstream drainage area and empirical relationships developed from sampled reaches. When FDISDAT.MDB is first imported into FHAT20, the total length of stream upstream for each reach is computed by summing the LENGTH field in REACH_CARDS. Total upstream length is strongly correlated with drainage area, which is a good predictor of some channel characteristics in areas of similar unit discharge (m^3/sec/km^2). The ratio of stream length to drainage area (km stream/km^2 drainage area) should be relatively consistent within an inventory area and will be a function of rainfall, surficial geology, and the detail that was used to represent stream lines on the TRIM maps.

Parameters of the power function predicting channel width, wetted width, or bankfull depth (Y) as a function of upstream length (UPLEN),

$$Y = a * UPLEN \wedge b$$

are estimated from the sample data (where widths and depths have been measured in the field) and applied to unsampled reaches. The model(s) is fit by a least squares procedure on log transformed variables.

Rather than estimate a single relationship for each variable (channel width or wetted width, etc.), FHAT20 allows the user to develop separate functions for different sets of reaches. For example, while channel width will be correlated positively with upstream length, we would expect unconfined reaches to be wider for a given upstream length than confined reaches. Hence it is logical to develop different relationships for subsets of the entire dataset, which we refer to as strata (because you are stratifying the data into different sets). Stratification can improve the accuracy and precision of the physical predictions assuming that there is a sufficient sample size to develop the empirical relationships within each strata. FHAT20 provides a mechanism to:

- define these strata based on attributes in the REACH_CARDS table of FDIS (remote-sensed attributes which are available for all reaches);
- build and evaluate models for each of these strata;
- estimate uncertainty in the predictions; and
- save the predictions to MODEL.MDB for display and use in fish range, habitat capability, and stream classification modelling.

To predict widths and bankfull depths, open the "Channel Morphology" dialogue box via the "Modelling" main menu item (Fig. 9). Select a variable to model (Channel Width = CW_avg, or Wetted Width = WW_avg, Bankfull Depth =BFD_avg) from the combo box at the top of the dialogue box labeled "y-axis." The relationship between this variable and upstream length for the sampled reaches will be shown in the adjacent x–y scatterplot as blue x's. Below the scatterplot is a white box containing a list of strata for the currently loaded stratification rule. If you click on a string in the white list box, the x–y scatterplot will display the data for the strata you selected. The

first item in the list box (0 / Unstratified) always shows the entire sampled dataset. If you want to load a different set of strata that had previously been defined, select one from the combo box labeled "Stratification." If you want to define a new set of strata, select the "Define Stratification Groups" choice below the "Modelling" main menu item (Section 3.2 below).
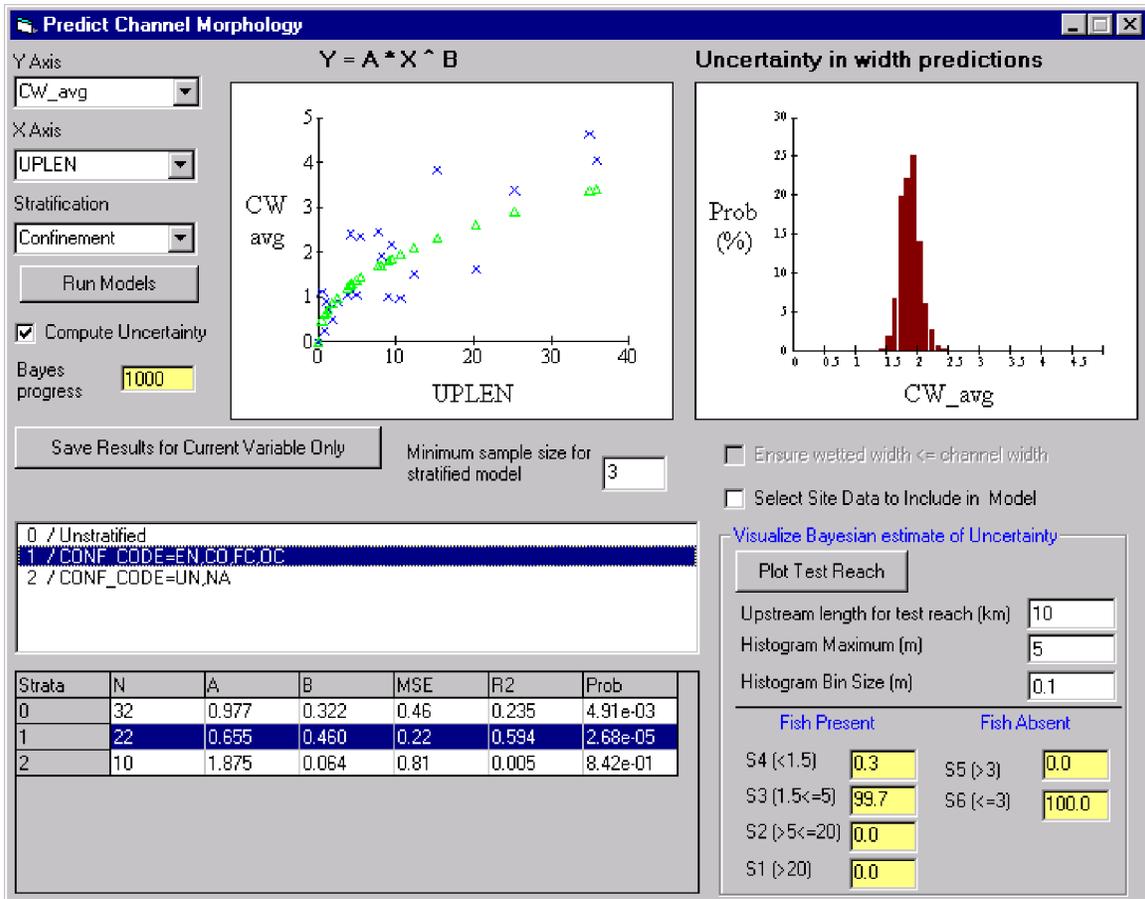
Figure 9: The channel morphology dialogue box is used to make reach-specific physical predictions, such as channel and wetted widths.

To fit power functions to the x–y data for each strata, click on the button labeled "Run Models." When you do this, model fit statistics will be displayed in the table at the bottom of the dialogue box and the fitted line (the model) will be shown as a green set of triangles in the x–y scatterplot. Fit statistics include sample size (N), constant (A) and slope (B) parameters of the power function, the unexplained mean square error (MSE), the correlation coefficient (R^2, the percent of the variance explained by the model), and the probability that the slope of the power function is not significantly different from zero (Prob.). Low MSE, high R^2, and low Prob. values denote good model fits to the data. You should evaluate these statistics across a range of stratification schemes to develop the most predictive models possible.

When examining how well the model fit a particular data set, you may notice outliers to the model, that is, blue points that are noticeably more distant from the fitted green line representing the model. These outliers may be normal sites that represent the extreme end of the uncertainty around the predictive relationship. Alternatively, you may know something about these sites which would motivate you to drop them from the model fitting because they do not reflect the

population of reaches you are trying to model. For example, there could be a problem with the measurements at a particular site (difficult to determine the top of the bank), or the site could be altered by human activity (e.g., a rip-rapped bank beside a road would not have a representative channel width). You would want to exclude these types of sites from the model fitting exercise as they could bias your predictions and levels of certainty. To exclude outliers, click on the check box labeled "Select Site Data to Include in Model." A grid will appear with the list of sites that are used in the model fitting (Fig. 10). Green cells denote sites that are included in the modelling, while red cells denote cells that have not been included. To include/exclude data from a site in the model fitting, toggle the color of the cell by clicking on it with the mouse. When you toggle a cell its corresponding reach will be highlighted on the map. Sites included in the modelling are recorded in the "UseFor__" ("_CW,""_WW,""_BFD") fields in the RES_PHYS table.

WS_CODE	REACH_ID	UPLEN	CW_avg	WW_avg	BFD_avg
120-907500	13		4.828571	3.5	0.4666667
120-907500	1	4.959	2.4	1.35	0.1
120-907500	7	29.37443	7.508334	6.016666	0.3033333
120-907500	1		0.9633334	1.075	0.13
120-907500	1	1.3355	4.891667	4.15	0.2833333
120-907500	1	3.9915	1.05	1.141667	0.4
120-907500	1	8.931	2.15	1.558333	0.2533333
120-907500	4	5.049	1.1	0.7916667	0.2166667
120-907500	4	15.144	0.9766667	0.985	0.18
120-907500	2	8.644001	0.9916666	1.288333	0.3833333
120-907500	2	23.294	1.616667	1.4	0.4
120-907500	1	0.4345	0.2533333	0.37	0.14
120-907500	1		1.045	1.241667	0.39
120-907500	4	11.3145	3.35	2.116667	0.3466667
120-907500	2	4.0135	0.895	1.011667	0.46
120-907500	2	1.656	0.7250001	0.7516667	0.2766667

Figure 10: The outlier dialogue box is used to display site-specific physical data used in channel morphology modelling, allowing users to exclude specific data from the modelling procedures.

Channel width is used in conjunction with fish presence to determine the FPC stream classification (S1–S6). Even with stratification, the empirical power functions you have fitted to the data will no doubt show substantial scatter. This means that there can be substantial uncertainty in the channel width predictions, and it would be unwise to only use the most likely (best-fit) width prediction in the decision making process. FHAT20 uses a bayesian algorithm (Walters and Ludwig, 1994; McAllister et al., 1994) to estimate the uncertainty in channel width for each unsampled reach. When you compute this uncertainty, the bayesian procedure essentially fits 1000 different power functions (different parameter values for A and B) to each data set (each strata). The relative likelihood of the data given each of these models (combinations of the A and B parameters) is computed and stored in memory. Once these 1000 models and their likelihoods have been computed, the procedure works its way through each unsampled reach and enters the total upstream length into each of these 1000 models to predict its channel width. A frequency distribution of predicted channel widths, generated by weighting each predicted width by the

likelihood of each of the 1000 models that was used to generate it, is computed for each reach. This distribution is then used to estimate the probability of each reach being in particular width classes (e.g., 1.5–5 m) as explained below.

To implement the bayesian algorithm, click on the check box labeled "Compute Uncertainty" and then click on the button labeled "Run Models" (Fig. 9). When the computations have finished, an example frequency histogram of widths will appear in the graphic adjacent to the x–y plot. This shows the uncertainty in channel widths for a theoretical reach with a known total upstream length. You can adjust the parameters effecting the display of the frequency distribution (e.g., the total upstream length, maximum of x-axis, bin size) in the frame labeled "Visualize Bayesian Estimates of Uncertainty in Width Predictions." If you change any of the parameters you must click on the button labeled "Plot Test Reach" to update the frequency distribution. Note that there will be different distributions for each model (strata) that was fit. The distributions will be relatively narrow (low uncertainty) for strata that have precise models (not much scatter), but will be relatively wide (high uncertainty) for imprecise models.

Below the frequency display parameters are a series of yellow boxes labeled S1–S6 (Fig. 9). When you have computed the uncertainty for a set of models, the values shown in these boxes display the probability that the test reach falls within each of the width classes associated with the 6 FPC stream riparian classification groups. This probability is simply the area under the frequency distribution within a specific width range (e.g., S2= 5–20 m). When you have computed the uncertainty for a set of models and save results to the RES_FISH table, a frequency distribution will be generated for each reach (it will not be displayed to save on computational time, but is generated internally), and the probability of the reach being in each of the six channel width classes will be computed. Eventually, when you compute stream classification, the probability of fish presence for each reach will be combined with its probability of being in each of the width classes to determine its FPC stream class designation. See section 8 for more details on stream class computations.

A few simple rules to remember when using the "Channel Morphology" dialogue box:
- If you want to save channel width predictions, you must first run the model(s) with the "Compute Uncertainty" box checked. When the computations have finished, click on the button labeled "Save Model Results to Res_Phys Table."
- If you want to save wetted width or bankfull depth predictions, you do not have to run the models with the uncertainty box checked (only the best-fit predictions for reach are saved to the database). Once the best- fit models predicting wetted width or bankfull depth have been computed for each strata, click on the button labeled "Save Model Results to Res_Phys Table."
- In cases where some of the stratified models have low sample sizes you might want to base predictions on the unstratified model for these reaches. Set the minimum sample size in the text box labeled 'Minimum Sample Size for Stratified Model' to define this limit.
- In some situations, the model may predict a wetted width that exceeds a channel width (based on stratified rules with few data points, or incorrect data, etc.). In such situations you may want to check the "Ensure wetted width ≤ channel width box prior to saving the results. The wetted width prediction for each reach will be compared to its predicted channel width and set to the channel latter value if wetted width exceeds channel width.

Note that you must predict channel and wetted widths for unsampled reaches as they are mandatory variables to perform the stream classification and habitat capability modelling steps, respectively (Fig. 7).

3.2   Stratification of Physical Predictions

Stratification groups are used to improve the precision of models predicting channel width, wetted width, bankfull depth, and the probability that a channel will be non-visible for reaches that were not sampled. A strata consists of a subset of reaches from the entire dataset defined by a set of remote-sensed characteristics. One or more variables can be used to define strata. For example, stream order could be used to define two strata, those reaches with stream order ≤ 2 and those with order >2. A more complicated stratification scheme or rule (also termed a stratification group) would be based on two or more variables, for example stream order and gradient. The previous stream order classes could be subdivided into 3 additional classes with gradients 0–2%, 2–5%, and >5% for a total of 6 strata. In the physical modelling, separate functions (predicting width and depth, probability of non-visible channels) will be fit for each strata. When saving model results, FHAT20 cycles through each unsampled reach in the dataset, determines its strata based on the variables you included in the stratification group, and then applies the appropriate model to predict its physical characteristics.

To define stratification groups, select the "Define Stratification Groups" choice from the "Modelling" main menu item (Fig. 11). Previously saved stratification schemes will be displayed in the dropdown box in the upper left corner of the dialogue box. When you select a scheme from this list, the strata groups will be displayed in the list box in the upper right hand corner.

To create a new stratification scheme, first click on the "Remove" button to remove any variables from the list box at the bottom left hand corner of the dialogue box (these were variables included in the currently selected stratification group). Then click on a variable from the list of available variables. A histogram will be displayed to the right of the list showing the distribution of values for the selected variable and the number of reaches from the total dataset with non-missing values. You do not want to select a variable to be included in the stratification scheme if there are many missing values. You can use the histogram to determine appropriate breakpoints for each strata class for this variable. To include a variable, select it and click on the "Add button." The variable will now appear in the list box at the bottom of the dialogue box. If you then select this variable again (from the lower list box) a list of unique values in the entire dataset will be presented to the right. You then need to define the number of bins (classes or breakpoints) for this variable. Click on a unique value and then on the text box for a particular breakpoint to populate that text box. You can also enter values manually. Note for string variables you can have multiple values for a single breakpoint. For example, if you used stream confinement code, you might define confined and unconfined classes. The former would consist of entrenched, confined, frequently confined, and occasionally confined reaches while the latter would consist only of unconfined reaches only. Repeat this procedure for other variables to include in the stratification scheme by adding additional variables to the lower-left list box. To save the new stratification scheme, type its name in the dropdown list box in the upper left of the dialogue box and click on the "Save Strata Group to Database" button. The stratification schemes will be saved to the ChanMorphStrata table in MODEL.MDB.
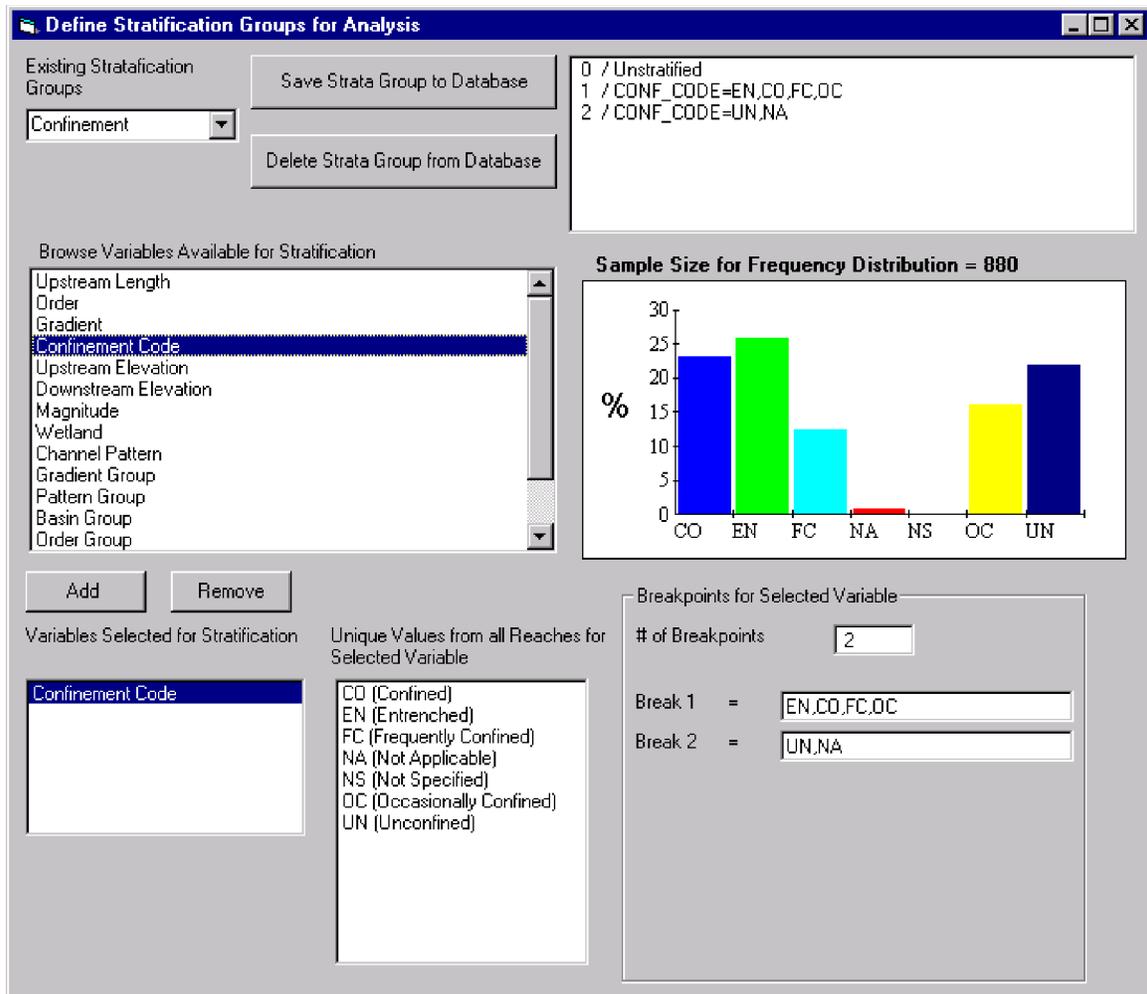
Figure 11: The stratification dialogue box is used to review and create rules that stratify data used to develop physical prediction models.

Keep in mind that there will be a strata class for all unique combinations of each variable-class that you define. For example, if you defined 3 stream order classes, two gradient classes, and two confinement classes, there would be 3*2*2=12 unique strata. These 12 strata will be used to develop separate predictive relationships for the physical modelling which has a sample size limited to the number of reaches that were sampled for physical data. If 36 sites were sampled, you would on average have only 3 sites per strata. More likely you would have some strata with 5–10 sites, and a number of strata with no or few sites. When you click on the combo box displaying existing strata groups, the number of sites in each strata class will be shown in the bar graph. You must trade-off possible increased precision obtained by more detailed stratification against the danger of fitting models with limited degrees of freedom (e.g., a linear regression based on two data points is fairly meaningless, even though its R^2 value will be 1). The model will default to using the unstratified predictive relationship for any reach that falls in a stratification class that has a relationship based on less than 3 data points (sites) or a larger value if you specify it in the Channel Morphology dialogue box. The stratification class for each unsampled reach for the last variable you saved in the physical predictions form (e.g., channel width) is saved with the predictions in the "WIDTH_STRATA" field in the RES_PHYS table.

3.3 Non-Visible Channel Predictions

Site assessment of some sampled reaches may reveal that the channel is not visible. This could signify that: 1) the mapping was incorrect and there was no channel where the map identified one; or 2) the channel may flow subsurface. In some watersheds, a substantial number of reaches could be non-visible channels. FHAT20 predicts the probability of each unsampled reach being a non-visible channel by stratifying the sampled reaches into different subsets (based on remote-sensed attributes that are available for all reaches – Section 3.2), and computing the percentage of sampled reaches in each strata that are non-visible. These probabilities are then applied to unsampled reaches in the same strata.

The "Non-Visible Channel" dialogue box, accessed via the "No Visible Channel" option from the "Modelling" main menu item is very straightforward to use (Fig. 12). A list box at the top of the dialogue box shows different stratification groups. When you select one, the total number of sampled *reaches* in each strata are displayed, along with the percentage of sampled reaches in this strata which were non-visible channels. You should examine these probabilities across different stratification schemes. The most predictive stratification is the one which provides the most contrast in probabilities among strata (i.e., some strata will have very high probabilities of a reach being a non-visible channel, while other strata will have low probabilities) while still maintaining sufficient degrees of freedom in each strata class. The model will default to using the unstratified predictive relationship for any reach that falls in a stratification class that has a relationship based on less than 3 data points. If an unsampled reach cannot be placed in one of the strata because it has a remote-sensed value that is outside the ranges of the stratification classes, the default (unstratified) relationship will be used to predict its probability of being a non-visible channel. The strata class for each unsampled reach is saved in the "NOVIS_STRATA" field in the RES_PHYS table.
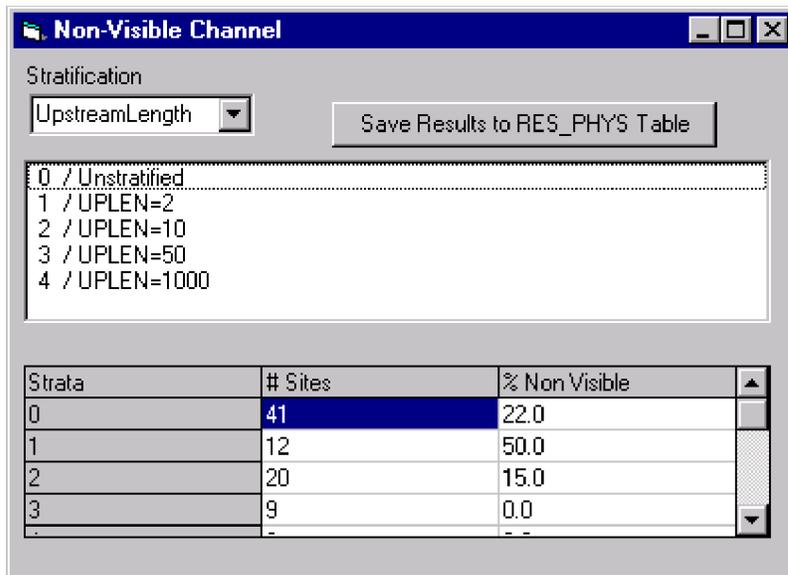
Figure 12: The non-visible channel dialogue box is used to predict the probability that unsampled reaches will be not be visible channels.

If you need to define a new stratification scheme, access the "Stratification" dialogue box from the "Define Stratification Groups" option from the "Modelling" main menu item (Section 3.2).

When you decide on the best stratification scheme, click on the button labeled "Save Results to RES_PHYS Table" to apply the model to the unsampled reaches and to save the results to the RES_PHYS table in MODEL.MDB.

3.4   Adding Physical Site Data to the Model Database

Additional physical data not contained in FDIS can easily be included in the FHAT20 model database to improve the precision and accuracy of predicted channel and wetted widths, bankfull depths, and the probability of non-visible channels. Open the USER_PHYS_SITE table in Access and enter the watershed code and reach ids for the new sites, as well as identifiers for the Site_ID. The latter field can be any numeric value (make one up if a Site_ID doesn't exist), and make sure to enter unique values for different sites in the same reach.

Enter any of the following information currently used in FHAT20 for physical modelling:

CW_avg = average channel width in m
WW_avg = average wetted width in m
BFD_avg = average bankfull depth in m
No_Channel_Vis = is the channel visible (Yes/No)

You can enter values for other data fields in USER_PHYS_SITE for completeness, but these data will not be used in any calculations in the current version of FHAT20. Since you presumably want to use these new data in the physical modelling, make sure the UseForCW, UseForWW, and UseForBFD fields are checked. The only circumstance when one of these fields should not be checked is if there are missing values for some of the measures (i.e., a channel and wetted width were measured, but no estimate of bankfull depth was taken).

You want to confirm that the KeepOnRebuild field is checked (default condition). This flag will ensure that the new data you enter will not be lost if you rebuild the database at a later date. Ensure that the FDIS_REC field is not checked (default condition). When user defined data is imported to the PHYS_SITE table, this field distinguishes FDIS data from user entered data.

After entering the data, close the USER_PHYS_SITE table and select the "Process Physical Site Data" sub-menu choice below the "Utilities" main menu item. This will initiate a procedure to ensure that there are corresponding records in the Reach_Cards table in MODEL.MDB for each record in USER_PHYS_SITE. If a user defined record has been successfully transferred to the PHYS_SITE table, the ValidSite field in USER_PHYS_SITE will be checked. If this field is not checked after you have "Processed the Physical Site Data," then the watershed code and reach_id for the site in question does not exist in the MODEL.MDB version of Reach_Cards, and ValidSite will not be checked. Edit the watershed code and reach_id in USER_PHYS_SITE so that it corresponds to a record in Reach_Cards and rerun the "Process Physical Site Data" procedure.

Note you will need to rerun the physical modelling procedures for these new data to have an effect on physical predictions.

4.0 Fish Groups

Fish groups are the basic unit used for modelling fish distribution and capability in FHAT20. A fish group is the sum total of all species and life-stages that make up that group. For example, the ONCORHYNCHUS fish group consists of all salmonids of the Oncorhynchus genus. Any reach where at least one fish belonging to this genus was caught or observed would be classified as present in status for the ONCORHYNCHUS group. The relative abundance of ONCORHYNCHUS In any reach, used for capability modelling, would be the sum of all fish belonging to this genus caught by electrofishing, divided by the site sample area. When an FDIS dataset is successfully imported into FHAT20, three fish groups are automatically created:

ALL_FISH: All fish species and age classes found in the watershed.
FPC_FISH: All species of fish found in the watershed used by the Forest Practices Code to identify a stream as fish bearing (Anon. 1998c).
ONCORHYNCHUS: All fish species and life-stages belonging to the Oncorhynchus genus.

You can create additional fish groupings via the Fish Grouping dialogue box. Note that a fish grouping can consist of a single species/life stage or a combination of different species and/or life-stages. This provides a very flexible framework for modelling specific species/life-stage combinations (e.g., 0+ rainbow trout) through to broad management groupings (e.g., all FPC species). At the minimum, you must model the range and capability of at least one fish group in order to predict FPC stream classes. When you create a new fish grouping you need to define which obstructions are barriers to migration and model its range and capability.

To define a new fish group, or examine what an existing fish group consists of, open the "Fish Groups" dialogue box accessed from the "Fish Groups" option below the "Modelling" main menu item (Fig. 13). If you click on any fish group in the dropdown list box labelled "Current Fish Groups" a text box at the bottom of the dialogue box will show the 'rule' (a SQL statement) that defines what species and life stages are included in the group.

- If you want to delete an existing fish group, select the group from the list box and click on the button labelled "Delete Selected Group."
- If you want to create a new fish group, type in the new fish group name in the dropdown list box, then:
    - Select the ages or life-stages to be included in the fish group. To add a particular age, life-stage, or species, select an entry from the one of the lists on the left side of the dialogue box and then click on the "Add" button and they will appear in the corresponding list box on the right side of the dialogue box;
    - If you want to remove one of the entries (an age, life-stage or species), select the entry from one of the right-side list boxes and then click on the button labelled "Remove;"
    - When you have defined the fish group, click on the button labelled "Process Records." Rules defining a fish group are saved in the GROUP_RULES table in MODEL.MDB.
- It is important to remember that if you add a new fish group you must go to the "Obstruction" dialogue box (section 5.0) and determine which entries in the obstruction table are migration barriers for this fish group before you can run fish range and capability models.
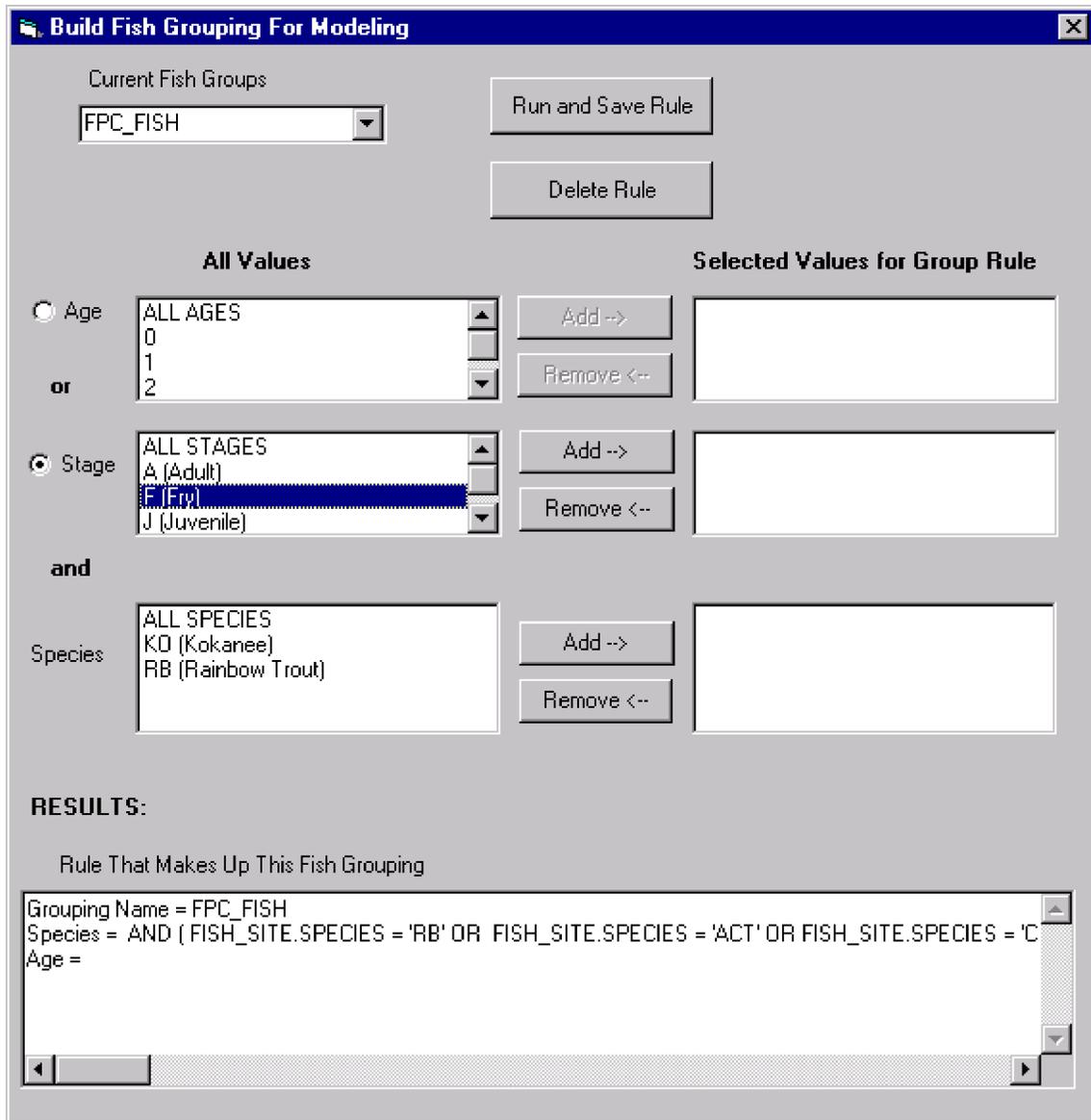
Figure 13: The fish group dialogue box is used to review and create new fish groups used in FHAT20 modelling procedures.

4.1 Adding Fish Site Data

Additional fish data not contained in FDIS can easily be included in the FHAT20 model database to improve the fish range, capability, and stream classification predictions. Open the USER_FISH_SITE table in Access and enter the watershed code and reach ids for the new sites, as well as identifiers for the Site_ID. The latter field can be any numeric value (make one up if a Site_ID doesn't exist), but make sure to enter unique values for different sites in the same reach.

The following fields must be populated:

Species:	Use FDIS codes.
Stage or Age:	Populate one or both of these fields using FDIS codes.
TOTALNO:	Total number of fish caught.
METH:	Type of gear employed, Use FDIS codes.
EFFORT:	# of seconds of electrofishing. Populate with missing value (-9999) if other gear type is employed.
AREA:	m^2 of area electrofished. Set to missing value for gear types other than electrofishing.
METHOD_NUMBER:	# of times a particular type of gear was deployed.

With the exception of the EFFORT and AREA fields, all other fields correspond to FDIS standards. Regarding METHOD_NUMBER, if 10 minnow traps were deployed at a site and only one species/age/stage was caught, there should be 10 records with METH='MT' with separate records for METHOD_NUMBER = 1–10. If two electrofishing locations were fished at a site, each location would be distinguished by a different METHOD_NUMBER. If more than one electrofishing pass was completed on a particular date, only enter data from the first pass (do not enter the average or the total across passes). If the site was resampled on a different date, enter separate records for each date (the data will be averaged when it is processed).

You want to ensure that the KeepOnRebuild field is checked. This flag will ensure that the new data you enter will not be lost if you rebuild the database at a later date. Enter a date in the Date field. Enter any comments describing the site or data source in the Comments field.
Ensure that the FDIS_REC field is not checked (default condition). When user defined data is imported to the FISH_SITE table, this field distinguishes FDIS data from user entered data.

After entering the data, close the USER_FISH_SITE table and select the "Process Fish Site Data" sub-menu choice below the "Utilities" main menu item. This will initiate a procedure to ensure that there are corresponding records in the Reach_Cards table in MODEL.MDB for each record in USER_FISH_SITE. The FHAT20 table summarizing the fish data by fish group (the MODEL_FISH_DATA table in MODEL.MDB) will then be rebuilt, and 'Network' type variables related to known fish occurrences will also be recalculated. Note you will need to rerun fish range, capability, and stream classification models after the new fish data have been processed. If a user defined record has been successfully transferred to the FISH_SITE table, the ValidSite field in USER_FISH_SITE will be checked. If this field is not checked after you have "Processed the Fish Site Data," then the watershed code and reach_id for the site(s) in question does not exist in the MODEL.MDB version of Reach_Cards. Edit the watershed code and reach_id in USER_FISH_SITE so that it corresponds to a record in Reach_Cards and rerun the "Process Fish Site Data" procedure.

5.0 Obstructions

Obstructions are used by FHAT20 to limit the distribution of a fish group in the watershed through the process of running a fish range rule (Section 6.0). All obstructions used in FHAT20 are stored in a table called USER_BARR in MODEL.MDB. When FDIS data are first imported into FHAT20, USER_BARR is populated with all the records from the FEATURE table in FDISDAT.MDB (thus it will contain features that are not obstructions, and you can delete these from the table if you wish). You can examine these records (i.e., the contents of USER_BARR) by opening the "Obstruction Editor" dialogue box accessed from the "Obstructions" option under the "Modelling" main menu item (Fig. 14). A table will show each feature, whose location is identified by a watershed code, reach_id. The feature 'measure' (distance upstream from the confluence) is also computed by assuming that the feature is located at the top of the reach it resides in (the measure value can be adjusted to reflect its actual position in the reach as described below).
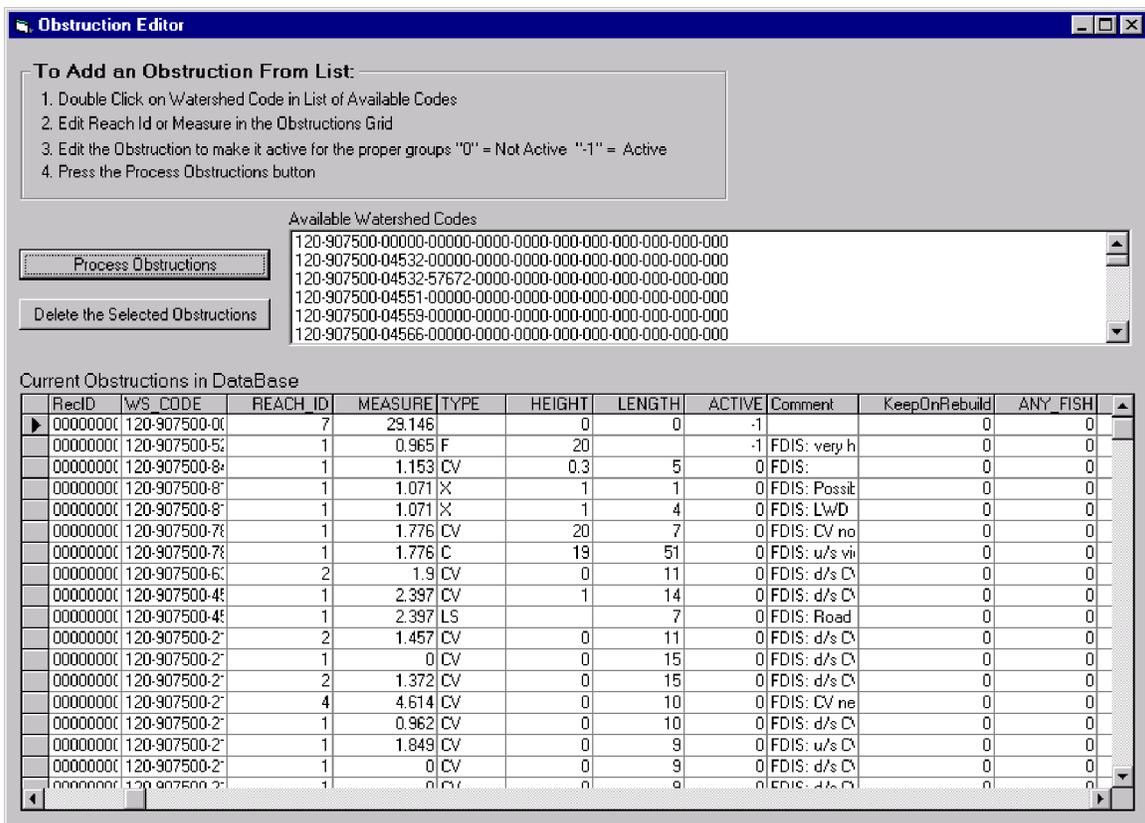
Figure 14: The obstruction editor dialogue box is used to display, edit, and add migration barriers to the FHAT20 model database.

Any characteristics of the feature, such as its type, height, length, and any comments that were entered in the FEATURE table will also be displayed. To view a map of the features listed in the table, select the "Features" theme from the "Physical Characteristics" table on the main form.

In the Obstructions table, the column labeled "ACTIVE" determines whether an obstruction has he potential to be a migration barrier for at least one fish group. The remaining columns on the right side of the table are labeled after the fish groups. TRUE/FALSE (-1/0) values denote

whether the feature is a migration barrier for that group. You must set these values for all fish groups for any obstruction that is active ("ACTIVE"=-1). **To navigate in the grid, click on the appropriate record and move to specific columns by using the left and right arrow keys.** Note the following conventions:

- If the ACTIVE field is set to 0 then the obstruction cannot be a migration barrier for any of the fish groups (e.g., FPC_FISH = FALSE).
- If the obstruction is a migration barrier for any of the fish groups, then the ACTIVE field must be set to –1.
- The ALL_FISH group represents all fish species. This fish group must therefore have the widest range in the watershed compared to any other fish group and would therefore be affected by the smallest number of obstructions relative to other groups. In other words, the ALL_FISH obstruction field should only be set to –1 if it is an obstruction to all species and life-stages found in the watershed.

There will be at least two circumstances in which you will want to add additional obstructions beyond those imported from FDIS:

1. You wish to include additional information on obstructions obtained from sources other than the reconnaissance survey (e.g., FISS information); and
2. A known upstream limit for a fish group that was not caused by an obstruction (e.g., temperature, channel morphology) cannot be modelled by the remote-sensed reach characteristics (gradient, order, width,…) used in the fish range rules. Assuming you have confirmed that there are no fish in this stream or reach by sampling, you can simulate this range limit by adding an 'obstruction', and then using the obstruction variable as part of the distribution rule (See Section 6.0). This allows you to exactly reproduce the known upstream limit of a fish groups range in a particular stream.

There are two ways of entering a new obstruction. The first is to click on a stream in the watershed code list above the obstruction table. When you do this you will be prompted as to whether you want to add this new record to the obstruction table. Once it has been added, you must specify the location of the obstruction in the stream. If only the reach_id is specified and the measure field (distance in km from the obstruction to the most downstream end of the stream (i.e., its confluence) is blank, the model will assume that the obstruction is located at the upstream end of the reach you specify and the 'measure' will be computed accordingly. Alternatively, you can enter the measure and the model will use this value as the barrier location. After processing the obstruction table, the reach_id corresponding to the measure will be shown in the table. Set the measure value to zero and the reach_id value to a legal entry if you want the model to use the upstream location of the specified reach as the obstruction position.

The second way of adding a new obstruction is to double-click on the desired obstruction location on the map (from any theme) in the main window. When you do this a "Search Results" dialogue box will appear with a list of reach attributes. If you have selected the correct reach to add the obstruction to, click on the button labeled "Add Obstruction to Selected Reach." The obstruction will then appear as a new entry at the top of the table and the reach_id column will automatically be populated. You then have the option of specifying a specific 'measure' for the obstruction, or to allow the model to assume that the obstruction is located at the upstream end of the reach. If you double click on a reach with an obstruction when viewing existing obstructions on the map (when viewing the "Features" theme), you can move directly to the appropriate row in the obstructions table by clicking on the "Select Record in Obstruction Grid" button on the "Search Results" dialogue box.

Once you have added a new record(s) to the obstruction table and defined the obstruction location via its measure or reach_id, you will need to fill out the other columns in the table. You can add notes about the obstruction, and the source of the information in the "Comments" column. Note that obstructions imported from the FEATURE table in FDIS are identified by the string 'FDIS' preceding the rest of the text in the COMMENT field in the obstructions table. Set the KeepOnRebuild field to True (-1) to ensure that the obstruction(s) you have entered will not be lost if you rebuild the model database at a later date.

Set the values for the "ACTIVE" column for new records to -1 (if you have added an obstruction, presumably it is an obstruction to at least one fish group). You must then set the values for all fish group columns. If you want to delete a feature, select if from the obstruction table and click on the button labeled "Delete the Selected Obstructions." When you are done editing the obstruction table click on the button labeled "Process Obstructions." This will calculate a field called "Group_DSBarr" (where Group = fish group name) in the NETWORK table in MODEL.MDB. This field is set to TRUE for all reaches upstream of any obstructions for this fish group (i.e., the obstruction is downstream of the reach in question), and FALSE for any reaches downstream of the obstructions. This new field can be viewed on the map by selecting a variable called "'Group' Downstream of Obstructions" from the Physical Variable Grouping. As a safeguard, you should always view a map of the obstruction results after you have edited and re-processed the obstruction table. It is very important that you include the "Upstream of Obstruction" variable when you build a fish range rule (Section 6.0) or the modelled distribution limits will likely be too widely distributed in the watershed.

In some cases, imprecision in the watershed codes may result in a tributary appearing to be upstream of a barrier in terms of how the model computes accessibility (when viewing the "Upstream of Obstructions" theme), even though on the hardcopy map the tributary is actually downstream of the barrier (and should therefore be accessible). If you are confident in the UTM value which determines the confluence location on the map for the stream in question, or know from fish sampling that the confluence is downstream of the barrier, then the barrier should be moved upstream by increasing its measure until it exceeds the confluence distance of the tributary(s) in question. This can easily be accomplished in a few iterations. Increase the value of the measure for the obstruction in question by x km, reprocess the obstructions, and view the "Upstream of Obstructions" field on the main form for the appropriate fish group. If the 'problem' tributary is still inaccessible, increase the measure again and repeat the process until the desired results are achieved (the tributary becomes accessible).

6.0 Modelling Fish Range in the Watershed

FHAT20 uses fish range rules based on remote-sensed reach attributes and obstruction data to model the potential range of a fish group within a watershed. Fish range represents the maximum potential distribution of the fish group in the watershed. This should not be confused with the computation of probability of presence in each reach, which combines information on known ranges, predicted potential ranges, and habitat capability. See Section 7.0 for details on how probability of presence is calculated.

A typical fish range rule might look something like the following:

> MAXDSGRADE<20 AND ORDER>=3 AND FPC_FISH_DSBARR=FALSE,

that states that the range of the FPC_FISH fish group will only be in reaches that: 1) have a maximum downstream grade less than 20% (the maximum gradient of any reach below it is less than 20%); 2) have a stream order greater or equal to 3; and 3) do not have any obstructions downstream (specific to the fish group, as defined in the "Obstruction" dialogue box). Because the range rules must be applied to all reaches in the watershed, only remote-sensed characteristics (from the REACH_CARDS table in FDIS) and predictions made for each reach in the watershed (excluding lakes, e.g., channel width) can be used as independent variables.

Note that range rules do not automatically produce a 'continuous' fish range within a watershed. Depending on what variables you use in the model, there can be reaches downstream of the upstream limit that are not included in the range. For example, the following rule,

> FPC_FISH_DSBARR=FALSE AND GRADIENT_20<15,

states that the FPC fish group must not have a barrier downstream specific to this group and cannot be present in reaches with gradients ≥15%. If there is a reach of gradient ≥15% downstream of a reach with gradient <15%, there will be a 'hole' in the fish range when viewing the results on the map (fish will be distributed upstream and downstream of a reach which is not included in the range). In the final computations of stream classification, probability of presence is adjusted so that a continuous fish range is used to classify the stream (Section 8.0).

Fish range in reaches classified as lakes in FDIS is predicted based on the predicted range value for the closest fluvial reach downstream of the lake. If fish are determined present in the first reach downstream of the lake, then the fish range is assumed to extend upstream into the lake (as long as the lake is within the modelled range limits). Alternatively, if the first fluvial reach downstream of a lake is predicted to be outside of the fish group's range, then the lake upstream is also assumed to be outside of this range. If you want to model the range of fish in lakes using this approach you must check the box labelled 'Predict range in lakes based on predicted range value in nearest downstream fluvial reach'. If you do not check this box, your range rule must include a component for lakes (e.g., OR isLake=TRUE), or be based on variables that are available for both fluvial reaches and lakes (e.g., MaxDSGrade, DSBarr, etc., but not channel width).

FHAT20 users can build a series of such range rules for any fish group, view the resulting ranges on the map, compare the predicted ranges with known distribution limits from the reconnaissance survey and other information sources (e.g., FISS), and modify the rules until the desired range is

achieved. If we knew, via sampling, the upstream range limits for a fish group throughout the watershed, there would be no need to go through such a modelling process. However, due to limitations on time and resources, this method is the only way to estimate or 'interpret' the range in parts of the watershed that have not been sampled. Note that by combining range rules with obstruction information, as in the examples given above, it is possible to perfectly replicate the range limits that are known while at the same time estimating the range limits in parts of the watershed that have not be sampled. The key to matching known range limits is entering them as 'obstructions' in the "Obstruction" dialogue box, regardless as to whether these observed limits are caused by actual migration barriers or other factors that cannot be modelled by the remote-sensed characteristics available in the reconnaissance dataset (e.g., temperature). One could also use obstructions to restrict the downstream range of a headwater-type fish groups or species. For example, a range rule with, 'Group'_DSBARR=TRUE would restrict the range of the fish group to reaches upstream of barriers for this group.

Review of existing fish range rules or entering new ones is done through the "Fish Range" dialogue box accessed via the "Fish Range" option under the "Modelling" main menu item (Fig. 15). Select the fish group to review or develop a range rule for from the combo box at the top of the dialogue box. If you have previously saved rules for this fish group, a list of them will appear in the combo box labeled "Rule Name." If you want to view one of these rules, simply click on the name and the rule will appear in the white text box at the bottom of the dialogue box. If you want to delete the rule click on the button labeled "Delete Rule." If you want to create a new rule, enter a new rule name in the combo box labeled "Rule Name." Delete the contents of the text box showing the last rule displayed. Then select a variable from the list box on the right side of the dialogue box. If you single-click on a variable in this list, the adjacent list boxes will be populated with:

1. unique values for all reaches in the dataset; and
2. unique values for reaches where the fish group was observed to be present.

The latter list will provide some guidance on maximum or minimum attribute values to use as part of the rule. Double-click on the variable to add it to the rule. The variable name (as it appears in the model database MODEL.MDB) will now appear in the lower-left text box. Select an operator (e.g., <, >, =,…) and a value from the list box of unique values (e.g., NETWORK.MAXDSGRADE < 20). Your rule will often contain more than one variable, so you will have to combine them with the 'OR' or 'AND' operators (see the example rule above). The 'LIKE' and 'NOT LIKE' operators are used to evaluate text variables such as confinement code (e.g., REACH_CARDS.CONF_CODE LIKE 'FC'). **All rules should contain the statement 'Group'_DSBARR=FALSE if there are known barriers for the fish group being modelled,** otherwise the modelled distribution could extend upstream of these obstructions.

After a fish range rule has been developed, it must be applied to all reaches in the dataset by clicking on the button labeled "Run Rule." As FHAT20 cycles through all the reaches in the dataset, it inputs the remote-sensed or predicted attributes for each reach into the model and determines whether the fish group is potentially present in the reach. This populates a field called 'Group'_DIST in the RES_FISH table in MODEL.MDB. If your rule is overly restrictive, the predicted range may not extend to reaches where the fish group was observed (false negatives). If your range rule is too liberal, the predicted fish range will extend into reaches that were sampled but where the fish group was not found (false positive). If you are confident that the fish sampling was accurate, that failure to find fish really meant that no fish ever use this reach, you may make the rule more restrictive. A list of false positive and false negative reaches will be shown in the dialogue box after a rule has been run. If you want to review the physical characteristics of these

reaches, simply double-click on them in the grids to highlight their location on the map. You then have the option of double-clicking on the actual reach on the map to see the attributes of interest.
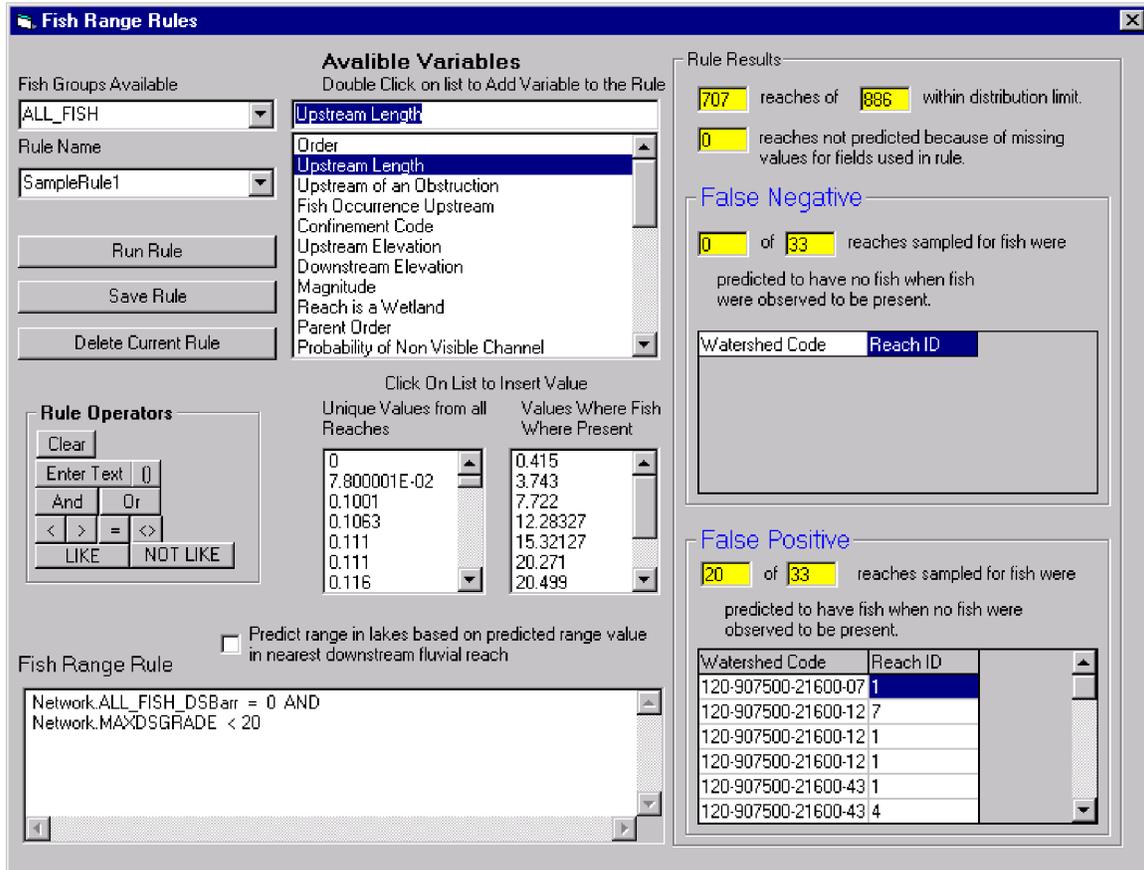
Figure 15: The fish range dialogue box is used to review and edit rules that predict fish group range in the modelled watershed.

Fish range results are automatically saved to the RES_FISH table in MODEL.MDB under the field name 'Group_DIST' (where Group represents the name of the fish group). Presence is denoted by a value of '1', absence by '0', and '-9999' denotes that a prediction could not be made because there was a missing value for at least one of the variables used in the fish range model. To save the rule itself to the RULES table in MODEL.MDB (the results are automatically saved when you run the rule), click on the button labeled "Save Rule." To view the results from a range rule, go to the main window and select the "Fish Predictions" variable group and the "'FISH Range" variable. To compare the modelled range with observed occurrences of the fish group, select the "'FISH Reach Presence/Abundance" variable from the "Observed Fish Data" variable group. Then click on the "Overlay Second Layer" check box on the main window and display the predicted fish range. The predicted range will appear as dots on top of the observed range. You will want to modify the range rules and possibly edit the obstruction table until: 1) the modelled range limits do not exceed the limits of observed occurrences if those occurrences are known to be upstream range limits; and 2) the modelled range limits extend, at a minimum, to all known occurrences. Remember that you can double-click on any reach on the map to look at its attributes. This will allow you to determine what variable in the rule has caused the range to stop at a particular reach. This information is helpful to modify the rule if required.

7.0 Modelling Fish Habitat Capability

Fish habitat capability is an index that measures the capability of stream reach to support fish. Capability can be assessed for a single species and life stage, or it can be an aggregate measure for all fish within the reach. Capability can rarely be measured in an absolute sense unless all fish within the reach can be caught. More likely, a consistent sampling approach is used to provide a relative index of fish abundance in various reaches and the relative 'capability' among reaches is then be compared. Estimation of absolute capability implicitly assumes that the reaches are 'fully seeded', i.e. abundance is not limited by juvenile recruitment to the reach. More realistically, a relative comparison of capability among reaches must assume that all such reaches are seeded to the same extent.

The Fish Habitat Capability component of FHAT20 uses fish density estimates (# fish/100 m^2 of sampled habitat) estimated by one pass-electrofishing as an index of fish habitat capability in sampled reaches. This assumes that catchability is the same in all reaches. The user builds various non-parametric models that relate remote-sensed reach attribute information such as stream width, order, and gradient to fish abundance estimates in reaches sampled by electrofishing. These models are then applied to unsampled reaches to predict their fish habitat capability using remote-sensed reach attribute information as input.

7.1 Overview of Method Used to Predict Fish Capability

The traditional approach to developing habitat capability models has been to rely on parametric, multivariate statistical tools, e.g. linear regressions that relate habitat attributes to an index of fish abundance. The use of these tools require that certain assumptions be met regarding the form of the habitat capability function and the distribution of errors among habitat attributes and across the range of abundance measurements. These assumptions however are rarely met (James and McCulloch, 1990; Rice, 1993) and as a consequence, tend to have weak predictive ability. Of even greater concern are the estimates of certainty about these predictions (Rice, 1993). Confidence interval calculations require even stronger adherence to modeling assumptions.

Consider the distribution of fish densities obtained by electrofishing across a number of sites with a range of habitat qualities (Fig. 16a) where in this example, habitat quality is indexed by water surface gradient (data from R. Ptolemy, BC Fisheries). It is clear from Figure 16a that habitat factors do not directly control the abundance of animals but rather, provide limits on maximum capability. In low gradient reaches, the majority of sites have low densities but a few sites are capable of high densities. As gradient increases, variability in densities is reduced and the high densities seen in reaches of low gradient are not attained. Gradient does not have much effect on the mean density but does limit the ability of a site to produce moderate to very high densities. If two 'slices' are taken through the bivariate distribution shown in Figure 16a this pattern can be seen more clearly (Fig. 16b). Note how the proportion of sites in the medium and high density classes (denoted by the vertical dashed lines) declines as reach gradient increases from <1% to 5–7%.

The functional forms of habitat-abundance data are certainly not simple. To paraphrase Rice (1993);

> *Although ecological theories can yield predictions of how animals should use habitats, theory predicts in only general ways the shape of specific abundance-habitat functions. When abundance and habitat data are plotted, the relationships commonly show combinations of thresholds, floor and ceiling effects, asymmetric ascending and descending limbs, marked skewness or kurtosis, differing variability in abundance at different positions along a habitat gradient and other diverse statistical problems. Curvilinear models may fit the data better than linear models, but they do not necessarily fit the data well.*

To overcome these problems, Rice (1993) suggested that a non-parametric density approach be used to predict abundance from habitat data. In particular, Rice (1993) advocates the use of kernel density estimation to predict or forecast probability distribution functions (pdf) of fish abundance for a given set of habitat attributes. A probability density function is simply a frequency histogram showing the probability (y-axis) of different fish densities classes (x-axis). The solid lines shown in Figure 16b are pdfs of fish abundance in high (gradient <1%) and low (gradient 5–7%) quality habitat. Perhaps the biggest advantage of the use of a pdf to estimate habitat capability in this application is that it does a good job of capturing the uncertainty in capability predictions.

FHAT20 uses multivariate kernel density estimation to compute the pdf of fish capability for every unsampled reach in the FDIS dataset based on a comparison of their attributes (e.g., gradient, order, width, elevation) relative to those in the sampled reaches where fish abundance estimates were measured. The simplest way of constructing a pdf for an unsampled reach would be to assemble density estimates from all sampled sites within the fish groups watershed range and plot this as a histogram. The kernel estimation method would simply draw a smoothed curve through this histogram (e.g., Fig. 16b) and we could compute various statistics of interest (e.g., median value, 95% confidence limits) from this distribution. If we employed this approach we are essentially saying that the capability of all reaches in the watershed are identical. Clearly, we can do better than this. We know intuitively for example, that abundance in a lower gradient stream tributary to a mainstem will tend to be higher for many species than abundance in a steep gradient stream in the headwaters of the watershed. So, if we were estimating a pdf for a low gradient reach, fish abundance data collected from reaches of similar characteristics should be given greater weight when estimating the pdf. Visualizing this graphically, the multivariate kernel estimation essentially constructs a surface predicting probability as a function of fish abundance and the habitat variable (Fig. 16a) based on all fish site abundance estimates (the reference data set), and then takes a vertical slice through this surface at the location of the reaches habitat value (Fig. 16b). This approach can be extended using multiple habitat variables but cannot be visualized graphically since we are limited to viewing results in three dimensions. See section 7.4 for computational and theoretical details of multivariate kernel density estimation.
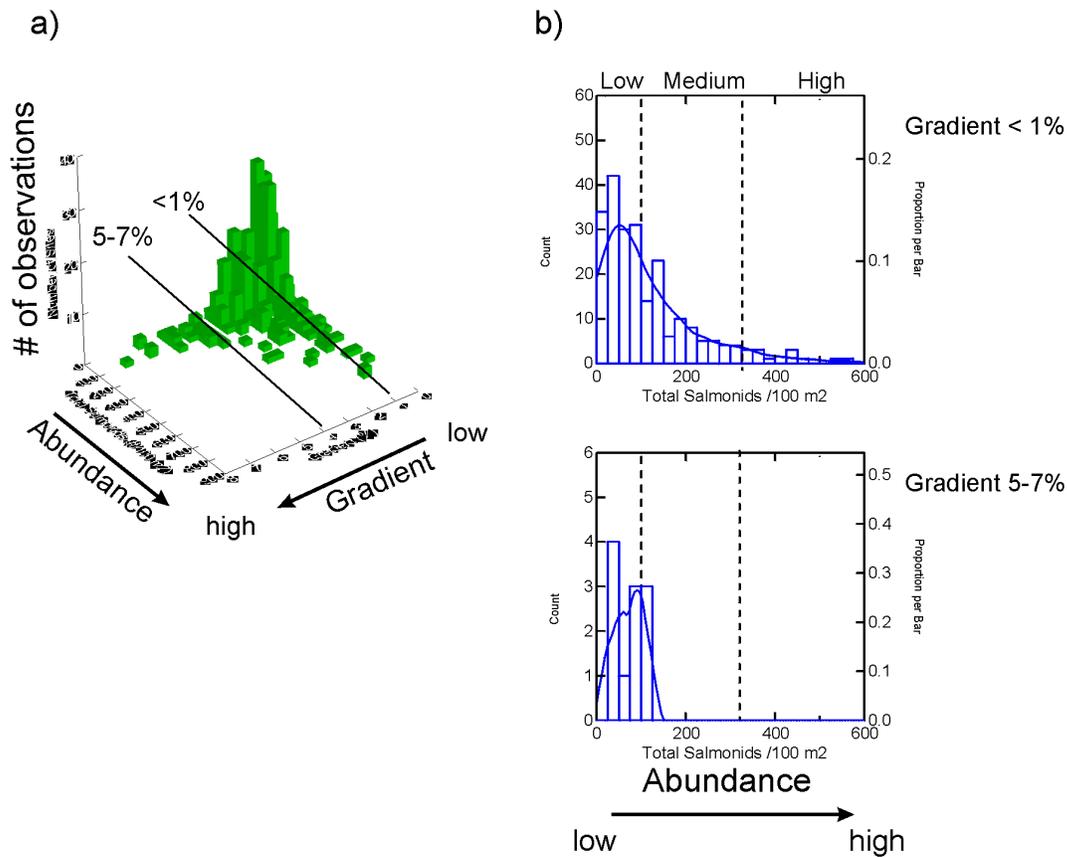
Figure 16: Relationship between habitat and fish abundance. Graph a) shows the bivariate distribution of total salmonids and stream gradient based on data from 77 coastal streams in British Columbia (data from R. Ptolemy, BC Fisheries). Graph b) shows probability density functions of total salmonid fitted to distributions taken from two habitat quality classes as indexed by stream gradient. Dashed lines denote low, medium, and high abundance classes.

The pdf estimation approach is ideally suited for data intensive situations such as the Reconnaissance (1–20,000) Fish and Fish Habitat inventory. Predictions and uncertainty are totally dependent on the input data. If there are strong relationships between habitat attributes and fish capability in the sample data, the kernel estimation will produce tighter (narrower) pdfs that are noticeably different between reaches with different habitat attributes. If the relationship between fish capability and habitat attributes is weak (often the case), then this will be reflected by wide pdfs (high uncertainty in capability) which look more or less the same in all reaches.

7.1.1 Prediction of Capability Classes based on a PDF

A probability density function is really just a plot of the probabilities of a range of fish densities for a particular reach (e.g., 0–600 fish/100 m^2 in Fig. 16b). This pdf must be summarized into simpler terms so that results for all reaches can be examined spatially on a map and summarized in tabular/statistical formats. One way of summarizing a pdf is to break the distribution into broader fish abundance classes and then estimate the probability for each of these broader classes. In FHAT20 we divide each pdf into low (<5 fish/100 m^2), medium (5–20 fish/100 m^2), and high (>20 fish/100 m^2) capability classes (classes can be redefined by the user). We essentially draw

such limits on top of the pdf (e.g., as vertical lines like in Fig. 16b) and then compute their probabilities as the integral of the area under the pdf curve within each class.

A pdf can also be used to compute the probability of fish absence, but this requires that a "not-present density limit" be defined. Intuitively, one would expect this density to be 0 fish/m^2, i.e., if the density is greater than this value, the site should be considered "Present" in status. This would indeed be the case if the entire area of the reach were sampled. In practice, sample site size is less than the entire reach area and in the case of a reconnaissance level inventory, sampled area is the product of the wetted width and 10 times the channel width or 100 m in length, whichever is greater. The minimum number of fish that can be caught within the site area to negate an "Not-Present" class status is obviously one (you can't physically catch less than one fish), thus the not-present density class limit must be 1 fish/site area. The 'not-present density' limit therefore decreases proportionally with the size of the site. Given two *reaches* with equal densities, one in a narrow channel (small sample site) and one in a wider channel (larger sample site), the "not-present density limit" (=1/site area) will be higher in the small site compared to the larger one. Assuming equal pdfs (densities) in these two sites for arguments sake, the probability of fish not-present will be therefore be lower in the larger site because we are sampling a larger area relative to the small site. FHAT20 accounts for this dynamic when computing the probability of not-present for each reach. Sample site size for each unsampled reach is estimated based on the predicted channel and wetted widths of the reach. The minimum density below which the reach can be declared to be 'Not-Present' of fish is 1 fish divided by this estimated site area. Large sites will have lower 'Not-Present' density limits than smaller sites.

A numerical example may help clarify these notions. Say you have two sites in two different reaches with equal densities of 1 fish/100 m^2 in both sites. One site is 100 m^2 and the other is 500m^2. Thus one site contains only one fish while the other has five. You classify a site as 'present' in status if one fish or more is caught, regardless of its size. Thus, under equal densities (1 fish/100 m^2 in this example), the probability of classifying the 500 m^2 site as not-present is 5 times less than the probability of classifying the 100 m^2 site as not-present. Thinking about it another way, assuming you sample both sites with the same degree of effort per unit area, you are five times more likely to catch a fish in the larger site since there are five times as many fish in your sampled area, even though the densities are the same. The bottom line is, that under equal densities, the probability of not-present is inversely proportional to the size of the sample site.

7.2   Using the Fish Habitat Capability Dialogue Box

The Fish Habitat Capability Dialogue box can be accessed from Fish Habitat Capability sub-menu choice below the "Modelling" main menu (Fig. 17). To compute pdfs of fish capability follow these steps:

1. **Select a fish group** from the combo box at the top of the dialogue box. By default, when the dialogue box loads, the currently active fish group is displayed. Note that before you can select a fish group to compute capability pdfs, you must first predict the fish group's range within the watershed via the "Fish Range" dialogue box. When you select a fish group, a reference data set (used to compute the pdf) is populated based on all sites sampled for fish within the fish group's watershed range. The number of sites included in this reference set is shown in a yellow box labeled "# of reference sample sites."

2. **Select habitat variables to include in the multivariate kernel estimation variables** from the table immediately below the fish group combo box. You can choose multiple habitat variables. To select the variable, click on the "In Model?" column to toggle the 'X' on (variable selected) or off. For each habitat variable, the table shows the habitat value for the reach that you will estimate the pdf for (see step 3 below), and the average, minimum, maximum, and standard deviation of values in the reference data set (sites that have been sampled for fish within the watershed range of the fish group). When you click on a row in the table, the graphic to the right of the table shows the relationship between this habitat variable and fish abundance in the reference data set (a two dimensional version of Fig. 16a). Each cell in the graphic shows the percentage of sample sites in a particular fish abundance – habitat value combination. The cells are color-coded (red = high percentage, light gray = low percentage) based on the legend shown at the top of the graphic labeled "% of sampled sites." The white text box to the right of the legend can be edited (followed by hitting the return or enter key) to rescale the legend. To determine which variables to include in the model, click on each row in the table and examine the values of cells in the adjacent graphic. Habitat variables that show strong patterns between abundance and the habitat attribute should be included in the pdf model.

3. **Select the reach to compute the pdf for**. The frame labeled "Process PDF for Single Reach" on the right side of the dialogue box contains a unique list of watershed codes. When you click on a watershed code in this list, the Reach_ID combo box below the list will be updated showing all reaches for this stream. Select the reach from this combo box to compute its pdf. An alternate way to select a reach is to close the "Capability" dialogue box to expose the map on the main window. Double-click on the desired reach. This will bring up a dialogue box showing the reaches attributes. If you close this dialogue box and open the "Capability" box again, you will note that the reach that you selected on the map will now be the currently selected reach in the "Capability" dialogue box. Its pdf will be displayed as well as its pdf statistics.

4. The pdf computed for a selected reach will be shown in the graphic in the lower left portion of the dialogue box. The pdf graphic shows the probability (0–100%) of different fish abundance classes for a continuous range of fish abundance classes (#/100 m^2). You can adjust the y- and x-axis maxima of the graph by editing the values in the text boxes below the graphic followed by the hitting the return key. The program automatically computes the probability of different fish abundance classes based on the pdf (i.e., it computes the area under the pdf curve within fixed ranges on the x-axis as in Fig. 16b). The fish densities that define each abundance class and their probabilities are shown in a frame immediately above the pdf graphic labelled "Fish Abundance Classes." If you want to alter the abundance class breakpoints (e.g., >20 fish/100 m^2 is the breakpoint for the 'High Abundance' class), simply edit the values in the appropriate text boxes and click on the button labelled "Recompute Abundance Classes." When you select a reach to process a pdf for, the estimated sample site size will be shown in the text box (labelled "Area of Sample Site") immediately above this button. This site area is the product of the estimated wetted width for the reach and 10 times the estimated channel width or 100 m in length, whichever is greater. Note that the density limit defining the "Not Present" abundance class is simply 1 fish/ Site Area (/100 m^2). As sample site area increases, the "Not Present" density limit decreases and so does the probability of fish absence.

5. Once you have examined the pdfs for a representative set of reaches and are comfortable with the model results, you are ready to process all reaches in the data set and save these results to the RES_FISH table in the model database. To do this, click on the button labeled "Process

PDFs for All Reaches" in the lower right hand corner of the dialogue box. This may take a few minutes depending on the number of reaches to process.
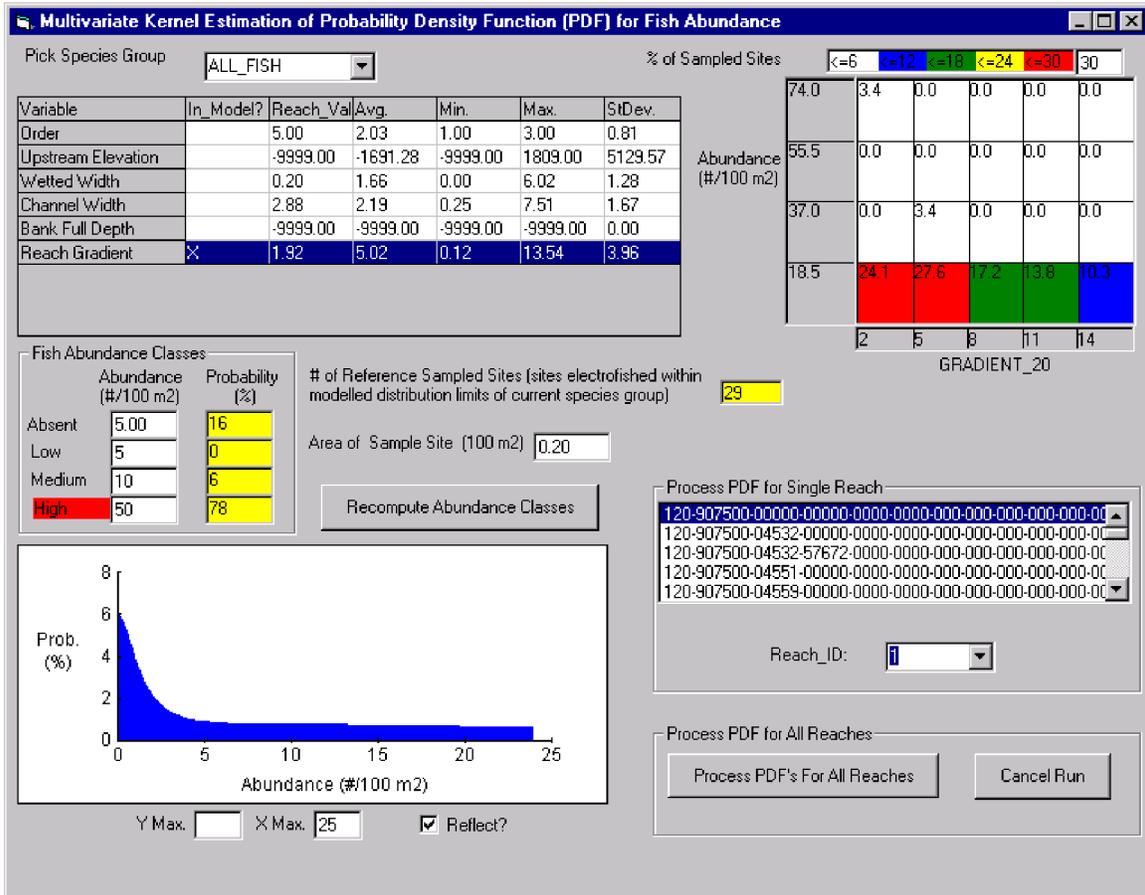
Figure 17: The fish habitat capability dialogue box is used to predict reach-specific habitat capability and probability of fish presence.

7.3   Output Indicators from Habitat Capability Modelling

There are seven output indicators that are produced by FHAT20 for each fish group that is modelled. All predictions are saved to the RES_FISH table by reach in the model database. This section describes what each of these fields represents. Results can be viewed spatially by loading the appropriate field from the "Fish Predictions" table in the main FHAT20 map window. Each field name is preceded by the fish group name (e.g., FPC_FISH, ALL_FISH), generically referred to as "Group" in the field descriptions below (See Table 3 to find the text descriptions shown in the main window which correspond to these field names).

Group_DIST: The maximum distribution (range) of the species within the watershed based on obstructions and user-defined distribution rules. '1' = potentially present, '0'= Not Present, '-9999'= could not be predicted because of missing data for input variable.

Group_Prob_Not Present, Low, Medium, High: The probability that fish will be not present in a reach or in low, medium, or high abundance classes based on habitat capability results.

Predictions are based on the pdf for unsampled reaches and sampled ones where fish were not found by electrofishing (or where electrofishing was not conducted). Note that the abundance classes represent habitat capability; if a site is above an obstruction it may still be given a high capability value (and low Prob_NotPresent value) even though it is currently not accessible (but could be if the obstruction was removed) and does not contain fish (see the Group_Prob_Presence field description). For sample sites where the fish group was found by electrofishing, the measured density is used to determine the appropriate capability class (set to 100%, all other classes set to 0%). For sampled sites were the fish group was found, but not by electrofishing, the pdf is used to estimate the probability of abundance classes but, the _Prob_Low class is the sum of the _Prob_Low value and whatever probability was estimated for the Not Present class (_Prob_NotPresent) and the Prob_NotPresent class is then set to zero (since fish were sampled in this reach, we know Prob_NotPresent must equal zero).

Group_Most_Probable_Class: The most likely capability class. This is computed by finding the maximum probability across the not present, low, medium, and high categories.

Group_Prob_Presence: Probability of fish group presence. This prediction combines habitat capability and fish range results. **Probability of presence is used in conjunction with channel width to predict FPC stream class (as described in Section 8.0 below).** If the reach is beyond the fish groups range in the watershed (Group_DIST=0) then this value is 0%. This applies to both stream and lake reaches. If the reach is within the group's range and downstream of a sampled reach where this fish group was found, then the value is 100%. If the reach is upstream of such a point, but within the fish group distribution range, then the value = 100 – Group_Prob_NotPresent as predicted by the pdf (Figs. 18 and 19). The only exceptions to this rule are for:

- Fluvial reaches that are inlets or outlets to lakes within the fish groups range. In this case, the probability of presence for the inlet/outlet reaches will be set to 100%.
- The first reach of a tributary flowing into a reach that is downstream of a known fish occurrence. In this case, the probability of presence for the tributary reach will be set to 100%.

Since a pdf is *not* computed for reaches classified as lakes, probability of presence is assumed to be 100% if the lake is within the fish group's distribution range, otherwise it is set to 0%.
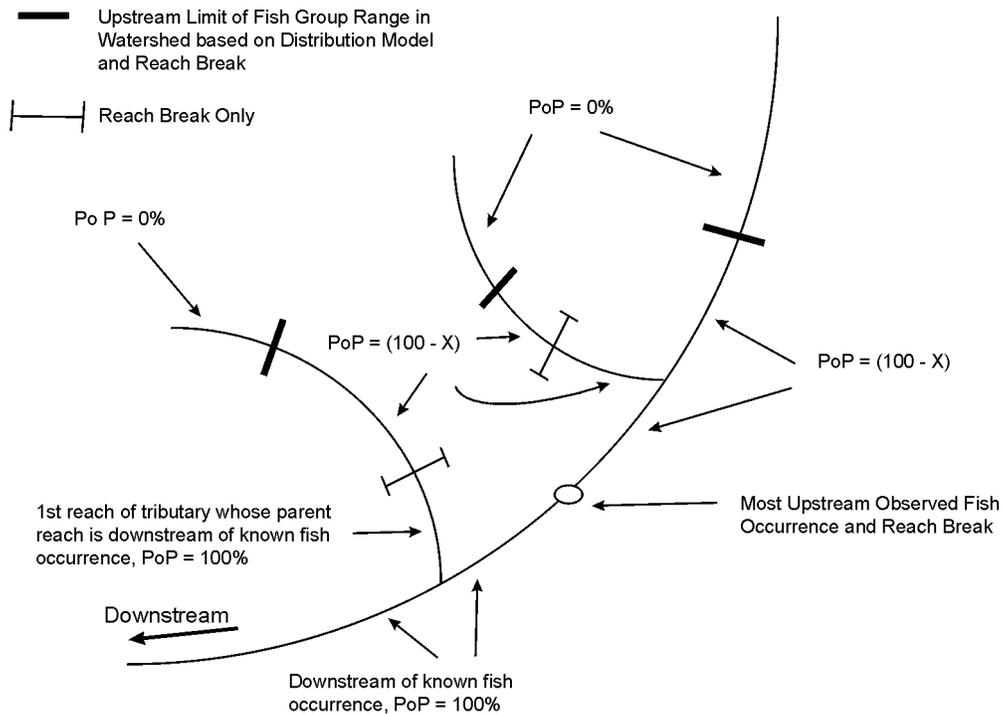
Figure 18: Schematic showing how the probability of fish presence (PoP) is computed for each reach as a function of modelled upstream distribution limits predicted by fish distribution rules and observed fish occurrences. If the reach is upstream of the modelled distribution limit then the probability of presence is set to 0%. If the reach is downstream of a known occurrence then the probability of presence is set to 100%. If the reach is upstream of a known occurrence but within the distribution limits of the fish group the probability of presence is generally 100 – X, where 'X' is the predicted probability that the capability in the reach is less than 1 fish over the typical sample site area for that reach (probability of not-present).
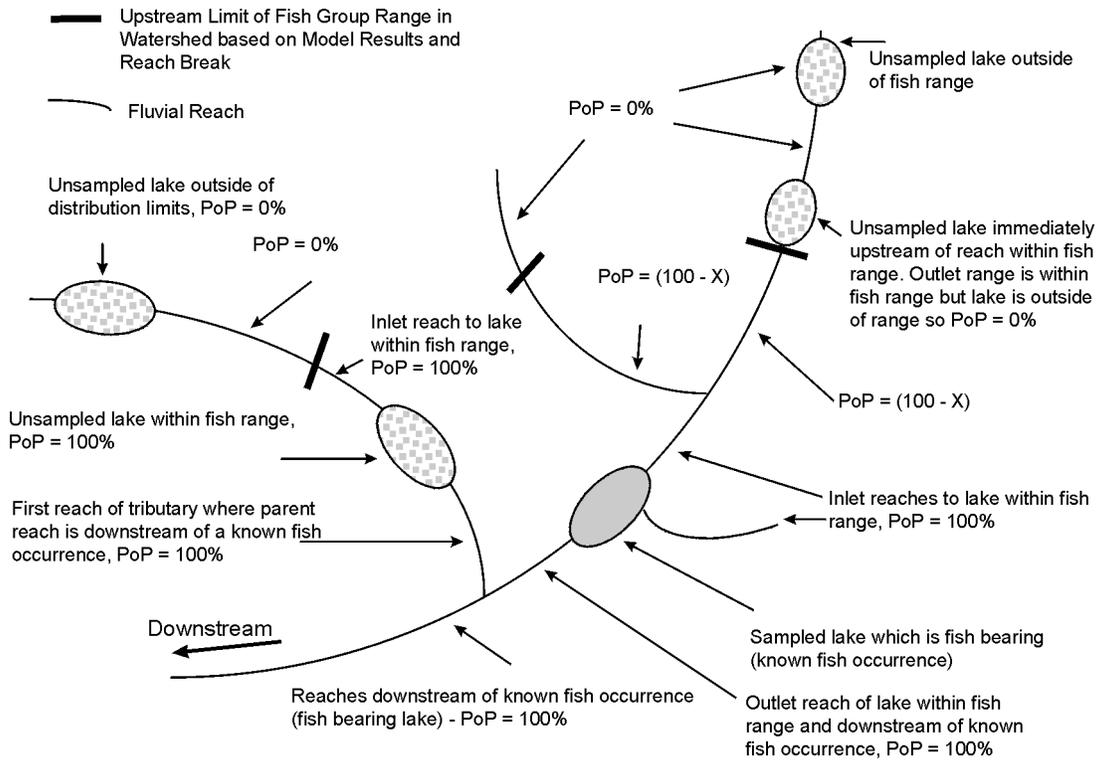
Figure 19: Schematic showing how the probability of fish presence (PoP) and fish distribution is computed for stream and lake reaches in relation to fish bearing lakes. 'X' is the predicted probability that the capability in the reach is less than 1 fish over the typical sample site area for that reach (probability of not-present).

8.0 FPC Stream Classification

All fluvial reaches in the model database can be classified according to the FPC stream classification system (Anon. 1998c). To perform a FPC stream classification, open the FPC Stream Classification dialogue box from the Stream Classification sub-menu choice under the Modelling main menu item (Fig. 20). Classification will depend on predicted and observed average channel widths combined with predicted and observed estimates of probability of fish presence. Thus the first step in classifying the reaches is to select a fish group to be used as the basis for determining fish bearing status. Since FHAT20 computes the fish group-specific probability of fish presence (PoP) for each reach on a continuous scale (0–100%, Section 7.3, Figs. 18–19), you must also specify a minimum probability of presence below which a reach will be classified as being non-fish bearing in status. Note that specifying a value of 0% (default condition) provides the most conservative predictions in that any PoP value >0% will lead to a fish-bearing classification for the reach. Once the minimum PoP value has been specified in the dialogue box, click on the "Compute Stream Classification" button to perform the calculations. Click on the "Save statistics…" button to copy the summary statistics from the stream classification (the table in the dialogue box) to the SclassStats table in MODEL.MDB.

FPC Stream Classification

Pick Species Group: ALL_FISH

Reaches with probability of presence less than 10 % will classified as non-fish bearing in status.

FPC Stream Class	# Reaches	Stream Length (Km)	Reserve Zone Area (Ha)	Management Zone Area (Ha)	Total Riparian Management Area (Ha)
S1	0	0.0	0.0	0.0	0.0
S2	17	20.6	123.6	82.4	206.0
S3	209	204.0	816.2	816.2	1632.3
S4	419	377.9	0.0	2267.7	2267.7
S5	2	1.5	0.0	9.1	9.1
S6	233	189.3	0.0	757.4	757.4
NVC	0	0.0	0.0	0.0	0.0
Lakes	6	0.0	NS	NS	NS
Totals	886	793.4	939.7	3932.7	4872.4

Compute Stream Classification   Save statistics to SClassStats

Figure 20: The stream classification dialogue box is used to classify reaches into S1–S6 FPC stream classes.

Results for the stream classification are presented in probabilistic and deterministic formats (Fig. 21). The probabilistic approach computes the probability of each FPC stream class (**Probability of S$_{1-6}$**) occurring in any reach, rather than generating a single most likely class.

This probability is generated by multiply the fish group probability of presence (PoP) for a reach by the probability of the appropriate width class for S1–S4 reaches, and 1-PoP times the probability of the appropriate width class for S5–S6 reaches. In all cases S1–S6 probabilities sum to 100%. The highest probability across the six stream classes is found and saved to the **Most Probable Stream Class** variable for each reach. For sampled reaches, the observed channel width is used to set one of the S1–S4, and S5–S6 width class probabilities to 100%, and all other probabilities are set to zero.

Uncertainty in stream class predictions for each reach is summarized in the **Uncertainty in S Class** variable. The variable compares differences in the probability of FPC stream classes (S1–6) within a reach. If there are fairly even probabilities across some or all of the classes, the uncertainty is high. The most uncertain situation is where all stream classes have the same probability (16.67% since there are 6 classes and the probabilities must sum to 100%); in this case the uncertainty in S Class variable will equal 100%. If one stream class contains a large probability (and others therefore have low probabilities), then the uncertainty is low. The most certain situation is where the probability for one stream class is 100%, and therefore 0% for the other classes; in this case the uncertainty value will be 0%. The formula for computing uncertainty is:

$$100 - \frac{\sum\limits_{i=1}^{n}\sum\limits_{i=1}^{n}|S_i - S_{i+1}|}{n-1}$$

where n=6 (6 FPC stream classes).

The deterministic estimate of stream class follows the FPC guideline that reaches cannot be declared non-fish bearing in status if they are downstream of fish-bearing ones. To reproduce this in the model, the stream classification procedure generates a **'Continuous' Probability of Presence** variable where PoP values remain constant or increase as one moves in a downstream direction (Fig. 22). A user-defined minimum probability of presence value is then overlaid on the 'Continuous' PoP map to generate a deterministic map of continuous fish presence, termed **FPC Fish Present**. The final stream classification, termed the **FPC Stream Class** is computed as either the most likely width class among S1–S4 ranges if the reach is fish-bearing in status (**FPC Fish Present** = True), or the most likely width class among S5–S6 ranges if the reach is not fish-bearing in status (**FPC Fish Present** = False). All results are saved to the RES_PHYS table, copied to the EXPORT table when results are exported, and can be viewed on the map.
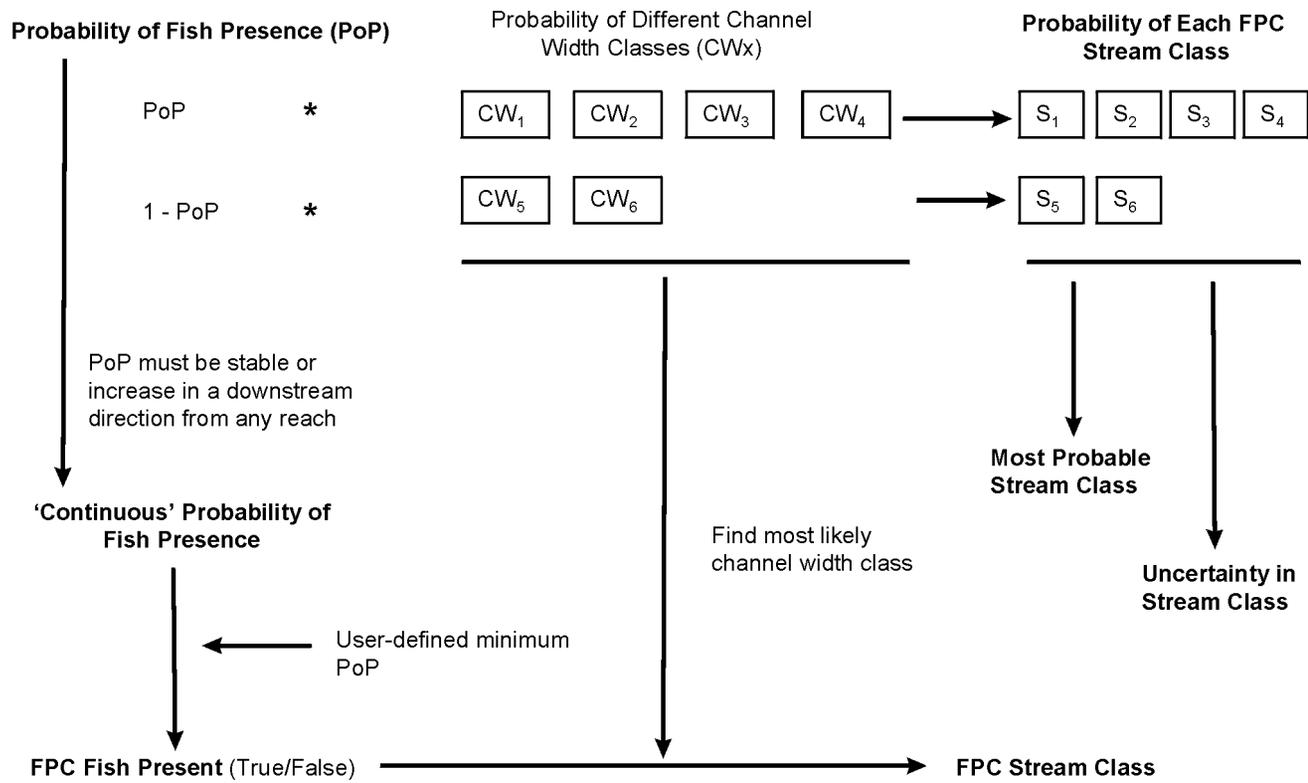
Figure 21: Schematic showing how stream classification predictions (highlighted in bold) are computed in FHAT20.
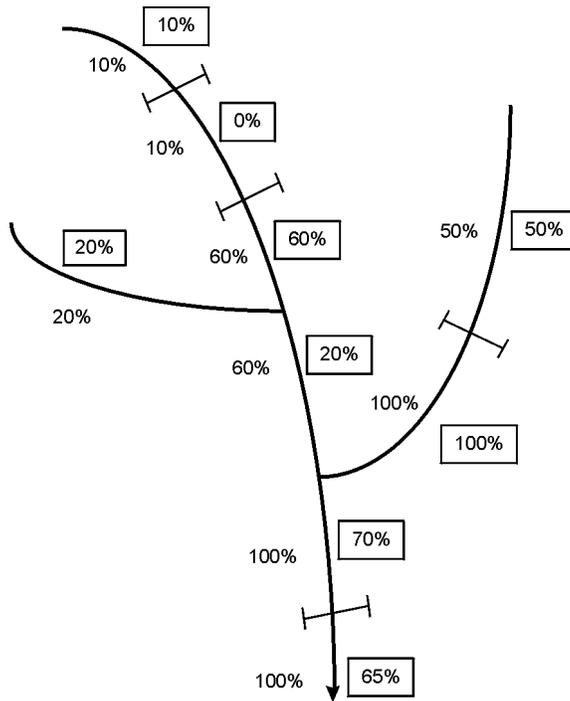
Figure 22: Schematic showing how continuous probability of fish presence (% values not enclosed in boxes) is computed from the probability of fish presence variable (% values enclosed in boxes) calculated in fish habitat capability model (Section 7.3, Figs. 18–19).

9.0 Editing Model Results

You may want to modify model predictions after closer inspection of the results. In some cases, this can be accomplished by simply re-running model rules for physical predictions, fish range, or fish habitat capability using different model variables and/or conditions. However, there may be cases where you are comfortable with the overall model(s) and only want to modify specific records where model predictions seem inappropriate. This can be accomplished by opening MODEL.MDB in Access and manually changing the values of specific records. Such 'manual' editing may be adequate when only a few records need be modified and when the database is relatively small. In other cases, an automated editing procedure, based on a series of user-defined 'editing rules', is required. For example, you may develop a fish range rule that limits fish to reaches with maximum downstream gradients<20% and to reaches with predicted channel widths > 3 m. However, from field sampling or experience in other systems, you may want your results to reflect the fact that many species can utilize the lowest reaches of small steep tributaries to larger mainstems where fish are present (i.e., confluence areas). You could edit the results using a rule that predicts potential fish presence if the reach is less than 3 m but is a tributary to a larger mainstem reach where fish were predicted to be present.

9.1 Automated Editing and the Edit Rule

Automated editing of model results can be accomplished via the "Edit Modelling Results" dialogue box accessed from the "Edit Results" option below the "Utilities" main menu item (Fig. 23). To edit a variable, select the variable from the list box on the left side of the dialogue box. If you want to edit a variable that is fish group-specific, you also need to select the appropriate fish group from the dropdown combo box where the fish groups are listed. When you click on the variable you want to edit, unique values will appear in the list box immediately to the right. You then need to select the independent variables you want to use in your editing rule. Select these from the right most list box. When you single-click on a variable in this list box its unique values will be shown in the middle list box. To include the variable in the 'editing rule' double click on it and it will appear in the text box at the bottom of the dialogue box. Set the appropriate conditions for that variable (e.g., Parent Order > 2) by typing the condition in the text box or by clicking on the math operators and double-clicking on the unique variable values in the middle list box. Finally, set the new value that you want to change the predictions to in the small text box above the large text box where you are creating the rule.

To run the edit rule, click on the "Run Edit Rule" button. Running the rule automatically overwrites any model results that meet the conditions of your rule. The number of records that are modified based on your rule will be reported. To save the rule to the model database, type a name for the rule in the combo box at the top of the dialogue box (labeled "Existing Edit Rules") and click on the "Save Edit Rule" button. To delete an existing rule, click on the "Delete Edit Rule" button.
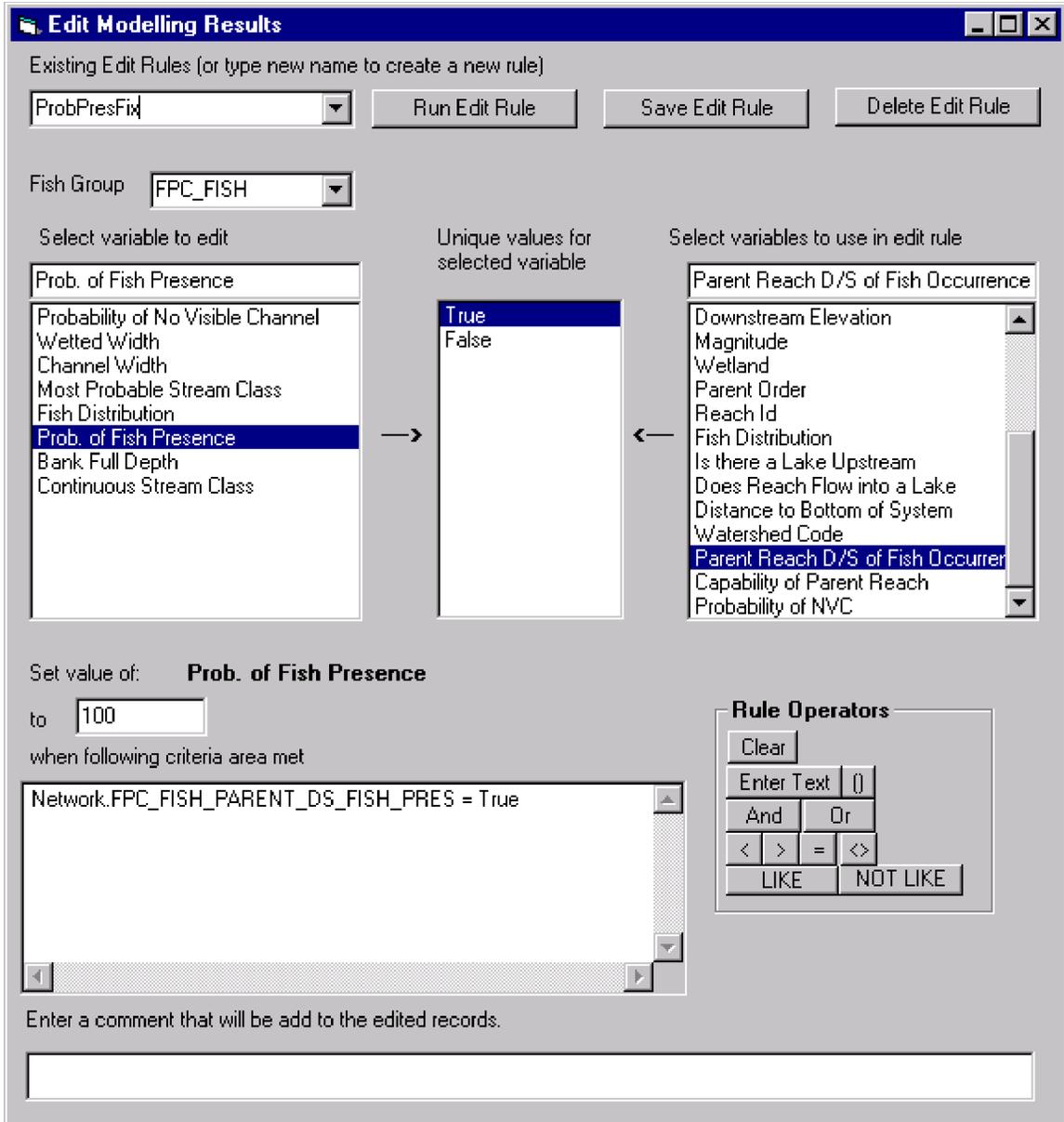
Figure 23: The edit results dialogue box is used to modify FHAT20 predictions based on a set of user-defined rules.

10.0 Controlling FHAT20 by Manipulating Tables and Files

Certain aspects of FHAT20 operation can be controlled by modification of tables in MODEL.MDB as well as the legend file MODEL.MDE.

10.1 Adding Variables to Include in Modelling

The DISTLOOKUP table controls what variables are available to use in various modelling procedures. The FieldName and TableName fields denote the name of the variable (used in the database) and the table that it resides in. The DisplayName field provides the descriptive string of the variable that appears in the various dialogue boxes used for modelling. The isGroup and isText fields denote whether the variable is tied to a fish group (e.g., FishGrp_DIST) and whether the variable is text or numeric, respectively. The remaining fields determine whether the variable will be available for use in various modelling procedures:

ForFishDist: variables can be used in fish distribution modelling (Section 6.0)
ForMKDE: variables can be used in fish habitat capability modelling (Section 7.0)
ForStrat: variables can be used in stratification for physical modelling (Section 3.2)
ForEdit: variables can be used for editing rules (Section 9.0)
EditableVars: variables can be edited (Section 9.0)

10.2 Controlling Data Checking

The DATACHECK table allows users to control how variables are screened for missing values during the importing process (see Section 2.2). The FieldName and TableName fields denote the name of the variable that will be checked and the table it resides in. The isText field denotes whether this variable is text or numeric in format. The remaining variables determine how the variable is screened and what happens when missing values are found:

MinLegalVal: the minimum value the variable can have (e.g., gradient ≥0%)
MaxLegalVal: the maximum value the variable can have
DefaultVal: the value that the variable will be set to in MODEL.MDB if it is outside of the range specified by MinLegalVal and MaxLegalVal
DeleteRecIfBad: determines whether a record with a missing value will be deleted if it is outside of the range specified by MinLegalVal and MaxLegalVal
IsActive: determines if the variable will be screened during the import process. This variable allows you to maintain the checking parameters in the DataCheck table but to not screen the variable (by setting isActive to false).

10.3 Adding Variables to Display on the Map

The ASCII text file MODEL.MDE, located in the project data sub-directory, controls what variables can be displayed on the map and how they are displayed (i.e., the legend). Modifying the legend (e.g., what colours are used, how many categories) for variables that can already be displayed is best accomplished through the FHAT20 Legend Editor (Section 2.3). However, if you want to display an additional variable from MODEL.MDB, you must make a new entry in MODEL.MDE for this variable. The easiest way to do this is to copy an entry for a similar type of variable (i.e., string, boolean, numeric) from MODEL.MDE to the bottom of the file, and then modify it. A typical entry would look like the following:

```
"Display Name","Order"
"Table Name","REACH_CARDS"
"Field Name","ORDER_20"
"Group Id",1
5,1,1,5
 16777215,1
 65535,2
 16711935,3
 255,4
 16776960,5
```

The first 3 rows of the entry are self-explanatory. The Group Id row determines which of the four mapping subsets (the uppermost combo box on the main FHAT20 form) the variable is associated with (Physical=1, Observed Fish Data,=2 Fish Predictions=3, Stream Classification=4). The next line has four fields: the number of categories in the display (e.g., 5), the type of variable (1=numeric, 0=text, -1=boolean), the lowest possible value in the dataset (e.g., 1), and the highest possible value (e.g., 5). The remaining entries specify the color for each category and the category number. Since the colors are represented by long integers, it is probably easier to modify the entry (what variable, number of bins, etc.) and then edit the colors from the Legend Editor in FHAT20 once you have made the changes.

Note to reload the legend file (MODEL.MDE) in FHAT20, you must close the program and re-open it. Also note that if you re-import the FDIS data any changes that have been made to MODEL.MDE will be overwritten, so it is probably a good idea to keep a copy of the modifications you make to MODEL.MDE in a separate file.

11.0   References

Anon. 1998a. Reconnaissance (1:20,000) fish and fish habitat inventory: Standards and procedures. B.C. Ministry of Fisheries, Fisheries Inventory Section for the Resource Inventory Committee.

Anon. 1998b. Reconnaissance (1:20,000) fish and fish habitat inventory: Data Forms and User Notes. B.C. Ministry of Fisheries, Fisheries Inventory Section for the Resource Inventory Committee.

Anon. 1998c. Forest Practices Code Fish-stream Identification Guidebook. Forest Practices Code of British Columbia Act, Operational Planning Regulation.

James, F.C., and C.E. McCulloch. 1990. Multivariate analysis in ecology and systematics: panacea or Pandora's box? Annu. Rev. Ecol. Syst. 21:129–166.

McAllister, M.K., Pikitch, E.K., A.E. Punt, and R. Hilborn. 1994. A bayesian approach to stock assessment and harvest decisions using the sampling/importance resampling algorithm. Can. J. Fish. Aquat. Sci. 51: 2673–2687.

Rice, J.C. 1993. Forecasting abundance from habitat measures using nonparametric density estimation methods. Can. J. Fish. Aquat. Sci. 50: 1690–1698.

Silverman, B.W. 1986. Density estimation for statistics and data analysis. Chapman and Hall, New York, N.Y.

Walters, C.J. and D. Ludwig. 1994. Calculation of bayes posterior probability distributions for key population parameters. Can. J. Fish. Aquat. Sci. 51: 713–722.

Appendix I: Theoretical Background of Multivariate Kernel Estimation

Kernel density estimation is analogous to constructing a frequency histogram. However, rather than assign counts to a set of predefined bins, a count is assigned to each observation in a data set that falls within a specified range (analogous to a bin) defined by a window width (x - h, x + h) and centred about the observation:

$$\hat{f}(x) = \frac{1}{2hn}\left[no.\,of\ X_1........., X_n\ falling\ in\ (x - h, x + h)\right] \tag{1}$$

The scalar (2hn) is used to transform the counts to estimates of probability such that the sum of all weights equals to one. Silverman (1982) refers to Equation (1) as a naïve estimator and is the simplest form of kernel density estimation. Kernel density estimation can be more formally defined as:

$$\hat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right) \tag{2}$$

where K represents the functional form of the kernel. In the case of the naïve estimator, the kernel function K is a simply a 'box' defined as:

$$K(t) = \begin{cases} 0.5 & if\,|t| < 1 \\ 0 \end{cases} \tag{3}$$

where,

$$t = \left(\frac{x - X_i}{h}\right) \tag{4}$$

and x = the observation of interest, Xi = all other observations in the data set. The naïve kernel estimator results in a pdf that is a discontinuous step function that does not meet the differentiability criteria of a true pdf. A much smoother pdf estimate can be obtained by considering other functional forms of the kernel. The most common and intensively studied kernel is the Gaussian (normal) distribution function where,

$$K(t) = \frac{1}{\sqrt{2\pi}} e^{-0.5t^2} \tag{5}$$

This is the kernel function used in the present algorithm. The resulting pdf estimate is continuous and integrates to one, thus meeting all the criteria of a true pdf.

One of the advantages of kernel density estimation over simple frequency histograms is that only one parameter must be set to construct a pdf. To construct frequency histograms, two parameters are required, the bin width and origin. Both parameters can have dramatic effects on the shape of

the resulting pdf estimate. Kernel estimates require only one parameter, the window width h, and will therefore always be a more robust estimate of the underlying pdf. The choice of window width h can be done subjectively through graphical assessments, based on a reference to a standard distribution, or done more objectively through cross-validation techniques. Variable window widths are also possible where h is varied depending on the local density of data. The use of variable window widths is a class of density estimation techniques that are referred to as adaptive kernels. To maximize computational speed, window widths in the present algorithm are calculated based on a reference Gaussian distribution. Window width in this case is analogous to the standard deviation of a normal distribution function and theory dictates that an optimum window width can be determined by a simple formula that considers the variance of the data set.

The discussion so far has been restricted to a univariate context where the objective of the estimation technique is to estimate the underlying pdf of a single variable. The kernel estimation techniques described above can be extended to include multiple variables with very little modification. The key difference is that each observation in the data set is viewed as a vector of variables rather than a single data point, where x and Xi are vectors of dimension d (corresponding to the number of variables). The multivariate Gaussian kernel function is as follows:

$$\hat{f}(\mathbf{x}) = \frac{1}{nh^d} \sum_{i=1}^{n} K\left(\frac{\mathbf{x} - \mathbf{X_i}}{h}\right) \tag{6}$$

$$K(\mathbf{t}) = (2\pi)^{-d/2} \exp(-0.5\, \mathbf{t^T t}) \tag{7}$$

$$\mathbf{t} = \left(\frac{\mathbf{x} - \mathbf{X_i}}{h}\right) \tag{8}$$

Conditional probability functions can be obtained by considering only the dimension of interest when calculating the pdf (in this case abundance). For example, in a two dimensional case (e.g., Fig. 16) where abundance and a single habitat variable form the vectors x and Xi, Equation. 6 can be used to construct a 3-D graph showing the probability (z axis) of a given abundance value (y axis) in relation to the habitat variable (x axis). A conditional pdf P(abundance|habitat value) can be obtained by taking a 'slice' through the 3-D relationship at the habitat value of interest (e.g., Fig. 16b). This is similar to calculating a univariate pdf, but using the multivariate kernel function (Eq. 6) with the habitat variable set to the value of interest. This rational can be extended to multiple habitat variables.

The use of conditional probability functions to develop estimates of abundance from habitat data is analogous to a non-parametric form of multiple regression analysis (Rice 1993). No a priori assumptions are made with respect to data distributions and functional form, thus avoiding the statistical difficulties of regression techniques in general. Further, predictive equations are not developed. Rather abundance predictions are in the form of pdfs that are constructed from the data itself.

The computational algorithms follow that described by Silverman (1982). All abundance and habitat variables are checked for symmetry using the g statistic for that describes the level of skewness. Variables that have g values greater than 1 are log10 transformed to restore a level of symmetry. The variables are then standardized to have an average of 0 and a standard deviation of 1. Transforming the variables in this way allows the window width to be calculated using a

simple relationship that has proven to be robust under a wide range of data conditions. Window width is calculated as follows:

$$Optimum\,Window\,Width = \{4/(2d+1)\}^{1/(d+4)}\,n^{-1/(d+4)} \tag{9}$$

A computational grid, on to which the pdf is mapped, is then created using the abundance data. Minimum and maximum values of the transformed abundance data are used to define a range of possible abundance values. The observed data range is then expanded by 4 times the window width to accommodate extrapolated values, and then divided into 150 equally spaced intervals. The result is a new grid variable that is used to map the pdf. A small scalar value is then computed and added to each grid point to ensure that at one of the grid points is equal to zero. A corresponding variable is also created to store the probability values that will be associated with each of the grid points as the kernel function is calculated.

The grid variable is then used to create a matrix of data consisting of the grid points and the habitat variables used as the predictor of abundance. The habitat variable values are identical for all of the grid point data. This matrix forms the set of x vectors used in Eq. 6. The transformed observations form the set of Xi vectors. To improve computational efficiency, all scalar calculations in the kernel function are moved outside the summation step. The resulting kernel density estimation function is as follows;

To improve computational speed, the summation component of Eq. 10 is done only for those grid points that lie within a ±5h range about the Xi of interest rather than the enter grid network. This results in a small loss in accuracy, but given the course resolution of the fish abundance categories, this loss is negligible. The scalar calculations only need to be calculated once.

$$\hat{f}(\mathbf{x}) = \frac{1}{(2\pi)^{d/2}\,nh^d} \sum_{i=1}^{n} \exp(-0.5\,\mathbf{t}^T\mathbf{t}) \tag{10}$$

Because the probability values will be extremely small numbers (10–16 – 10 –32), there is a high risk of numerical error when scaling the pdf using the appropriate scalar function. Another error is introduced because of the limited range of the grid, which cut off the very extreme ends of the tails. The cut off tails are not significant in practical terms as they tend to be extremely small values (< 10–32), but they never the less affect the numerical integration of the pdf at the computational level. Another problem is that the scalar formula applies to the full multivariate pdf, and not the conditional pdf of interest. To avoid these complications, the scalar is estimated by integrating the pdf estimate as calculated in Eq. 9 and then dividing the probability of each grid point by the integrand. The result is a pdf that integrates fully to one with a 10–10 order of magnitude for accuracy.

Because the gaussian kernel is not bounded in any way, negative abundance values will be possible. There are two ways to deal with these negative abundance data. The first is to simply accept the negative values and include their integrand as part of Fish Absence probability estimate. From a numerical perspective, the disadvantage of this approach is that the pdf on the positive line will not integrate to one and therefore will not be considered a true pdf. Further, the resulting pdf may not reflect reality. The fact that no fish are caught at a site does not eliminate the possibility that there may indeed be fish in the reach (at a location or time other than the sample site and date) and reflects a limitation in the sampling procedure. If one were absolutely

certain that fish were absent from a reach, then this approach may be appropriate. But this certainty is not possible given the sampling procedure. A more conservative approach that accounts for this sampling bias is to 'reflect' the probability values associated with negative abundance values onto the positive line and then ignore all negative abundance values. This approach is the preferred option suggested by the Silverman (1982) and is the default option adopted in the present algorithm. In the Fish Capability dialogue box, you have the choice to reflect –'ve values by checking the box labelled "Reflect?. The main drawback of this approach however, is that the probability of fish absence may be underestimated if, in sampling, we are indeed certain that fish are absent from a given site. It will be useful to empirically evaluate both approaches by comparing their respective predictive capabilities.

Once the data are reflected, the grid points are back transformed and the associated probabilities are integrated (summed) across each of the abundance categories as noted above. The result is a probability value that one of the four abundance categories is appropriate for a given site.