Productive Capacity of Great Lakes Nearshore Fish Habitats: Design and Implementation of an Electrofishing Survey Database

J.E. Moore[1], C.K. Minns[2], B. Valere[2], and R.G. Randall[2]

[1] JEMSys Software Systems Inc.
22 Marion Crescent
Dundas, Ontario L9H 1J1

[2] Great Lakes Laboratory for Fisheries and Aquatic Sciences,
Department of Fisheries and Oceans
Bayfield Institute, 867 Lakeshore Road, PO Box 5050
Burlington, Ontario L7R 4A6

July, 1998

Canadian Manuscript Report Of Fisheries And Aquatic Sciences No. 2441

Fisheries and Oceans
Pêches et Océans

Canada

Canadian Manuscript Report
of Fisheries and Aquatic Sciences 2441

July, 1998

Productive Capacity of Great Lakes Nearshore Fish Habitats: Design and Implementation of an Electrofishing Survey Database

by

J.E. Moore[1], C.K. Minns[2], B. Valere[2], and R.G. Randall[2]

[1] JEMSys Software Systems Inc.
22 Marion Crescent
Dundas, Ontario L9H 1J1

[2] Department of Fisheries and Oceans, Great Lakes Laboratory for Fisheries and Aquatic Sciences, Bayfield Institute, 867 Lakeshore Road, PO Box 5050
Burlington, Ontario L7R 4A6 Canada

Cat. No. FS97-4/2441 ISSN 0706-6473

Correct citation of this publication:

J.E. Moore, C.K. Minns, B. Valere, and R.G. Randall. 1998. Productive Capacity of Great Lakes Nearshore Fish Habitats: Design and Implementation of an Electrofishing Survey Database. Can. MS Rep. Fish. Aquat. Sci. 2441:vi+32p.

TABLE OF CONTENTS

ABSTRACT

J.E. Moore, C.K. Minns, B. Valere, and R.G. Randall. 1998. Productive Capacity of Great Lakes Nearshore Fish Habitats: Design and Implementation of an Electrofishing Survey Database. Can. MS Rep. Fish. Aquat. Sci. 2441:vi+32p.

A database was established to hold and organize the large amount of electrofishing data collected for the Productive Capacity project from 1988 to 1997. The Productive Capacity project was implemented to collect fish and habitat data from the field, and to determine linkages between fish assemblages and habitat in different nearshore areas of the Great Lakes (the lower lakes and Georgian Bay). This report provides details of the database structure, including hardware and software requirements, the database design (description of the table summaries and table relationships), bulk loading procedures, data retrieval features, and methods for dealing with missing data (fish weights). The database currently contains 1634 samples and 45,392 individual fish records from data collected at 17 different nearshore locations in Lake Ontario, Lake Erie and Georgian Bay.

RÉSUMÉ

J.E. Moore, C.K. Minns, B. Valere, and R.G. Randall. 1998. Productive Capacity of Great Lakes Nearshore Fish Habitats: Design and Implementation of an Electrofishing Survey Database. Rapport manuscrit canadien des sciences halieutiques et aquatiques 2441:vi+32p.

Une base de données a été montée en vue d'archiver et d'organiser la grande quantité de données sur la pêche électrique qui ont été recueillies entre 1988 et 1996 dans le cadre du projet sur la capacité de production. Ce projet avait pour but de recueillir sur le terrain des données concernant les poissons et leur habitat, et de définir les liens entre les assemblages de poissons et les habitats dans différentes zones littorales des Grands Lacs (lacs d'aval et baie Georgienne). Le rapport donne des détails sur la structure de la base, avec ses exigences en matériel et en logiciel, sur sa conception (description des sommaires des tableaux et des relations entre eux), sur les procédures de chargement en vrac, sur les caractéristiques de l'extraction des données et sur les méthodes à suivre pour les données manquantes (poids des poissons). À l'heure actuelle, la base contient des données sur 1 634 échantillons et 45 392 fiches sur des poissons, compilées à partir des données recueillies à 17 stations littorales du lac Ontario, du lac Érié et de la baie Georgienne.

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

INTRODUCTION

The purpose of this report is to describe the design and implementation of a relational database holding the cumulative data records gathered in electrofishing surveys at a number of locations in the nearshore waters of the Great Lakes.

The electrofishing database was established to support the Productive Capacity project at the Great Lakes Laboratory for Fisheries and Aquatic Sciences (GLLFAS), Burlington, Ontario. The general objectives of the Productive Capacity project were to conduct detailed fish and habitat surveys in the littoral zone of the Great Lakes (lower Lakes and Georgian Bay) to determine the fish community composition, species richness and biomass in areas with different habitat attributes, and to define linkages between fish and habitat in the near shore zone. Valere (1996) documented the locations, sampling chronology, and methods used to conduct the fish surveys between 1988 and 1995. Analysis of the fish and habitat data have been documented in a number of reports, including Minns et al. (1993, 1994); MacLeod et al. (1995); and Randall et al. (1993, 1996). Further data analysis is currently ongoing.

This report provides the following details of the database: hardware and software requirements, the database design (description of the table summaries and table relationships), bulk loading procedures, data retrieval features, methods for dealing with missing data (fish weights), and a summary of the current database holdings.

SYSTEM OVERVIEW

History

The large amount of electrofishing data collected for the Productive Capacity project (and others) was originally held in the form of spreadsheet-like tables within a desktop computing package called RS/1 (formerly BBN, now Domain Solutions). The data were first gathered together in a relational database format as part of an Oracle database developed for GLLFAS by Moneco Inc. in 1993. That environment provided enhanced security with respect to loss and corruption of the data and a much enhanced ability for reporting on and extracting the data. Unfortunately, the services of the database administrators were subsequently lost and there existed no means to add new data to the database.

To address that problem, it was decided to accept some loss of security and to port the database back to the desktop computer into a relational database format providing the organizational and reporting features obtained under Oracle, but in an environment providing more immediate control of the database than was possible under Oracle. The database chosen for this purpose was Microsoft Access 2.0, because of its relative ease of use and because it would be relatively easy to write interface software that would ultimately allow the whole database to stand alone.

The Oracle database table structures were imported into Access, a front-end which included a bulk-loader was written, and the text files originally used to load the Oracle database were used to populate the tables in the new Access database. Queries were written to support the desired extraction reports, and additional software was written to implement those extractions in a flexible manner.

Hardware And Software

The electrofishing database package consists of two parts:

1. The database itself is contained in a single Microsoft Access database (MDB) file called ELECFISH.MDB, generated using the 16-bit Microsoft Jet 2.5 database engine. This file may reside on the local PC or network. Access itself is not required to deal with the database.
2. Software was originally written in Microsoft Visual Basic 3.0 and created for the purpose of doing data input, updates, and extractions. That program was later ported to Visual Basic 4.0, but remains a 16-bit application.

Both the software and database may be installed on any i386-compatible computer running Windows 3.x or Windows 95, and should also run under Windows NT 3.x or 4.0, although this has not been tested. A practical lower limit on hardware would be a 486-33 machine with 8M RAM and 10M free hard disk space.

DATABASE DESIGN

Database Table Summary

At present the database consists of eleven tables, listed and summarized in Table 1. A complete description of all tables is provided in Appendix A.

Table 1. Summary of ELECFISH database tables and their contents.

Table:	Each row has information about:
LOCATION	One geographic location or area within which sampling stations are located
STATION	One station or point at which one or more samples were collected, e.g., the centre-point of an electrofishing transect.
SAMPLES	One sampling event at which time one or more species of fish were collected
ELECTRIC_FISHING	Electroshocking duration and operating conditions during a sampling event

Table:	Each row has information about:
WATER_OBSERVATION	Related observations (such as temperature and Secchi depth) during a sample event
TAXON	The taxonomy of one fish species as identified by its unique ACCESS_CODE
FISH_GROUP	Weights and counts for one species captured during a sampling event
FISH_INDIVIDUAL	Weight, length, and other data for one individual fish captured as part of one group
CODEFLD	The meaning of a code present in the database
Updates	Audit information related to one bulk-load event during which a batch of rows was loaded into one database table
Regression Terms	Length-weight regression sums-of-squares information for one species pooled over samples collected at one location

The first eight of these tables are the principal tables in the database, i.e., they contain data or information about data collected in the field. The relationships of this principal group of tables to one another are shown in Figure 1. Note that all relationships depicted are enforced by referential integrity constraints with cascading updates and deletes, meaning, for example, that if a row should be deleted from the SAMPLES table, all rows in the ELECTRIC_FISHING, WATER_OBSERVATION, FISH_GROUP, and FISH_INDIVIDUAL tables related to that sample would also be deleted. In general, therefore, it is not possible to strand rows in child tables by deleting (or updating) rows in parent tables.

This feature can be exploited to provide a means of reversing the effects of a bulk-load. Every row in all database tables has a batch update identifier which points to a row in the Updates table. Though not depicted in Figure 1, this table is a parent for every other table in the database, with cascading deletes enforced. Should a row be deleted from the Updates table, all rows loaded during the corresponding bulk-load event will be deleted from the database.

Database Table Relationships

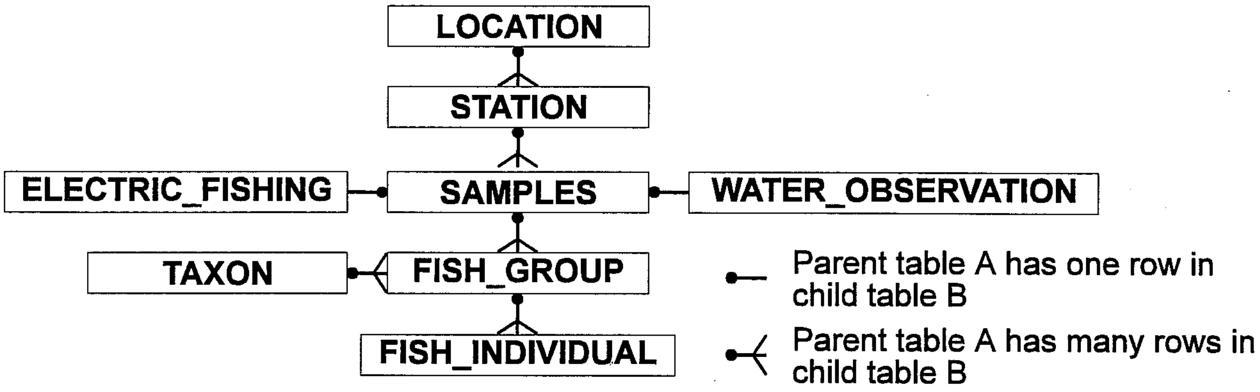


Figure 1. Entity relationships diagram for the principal database tables.

BULK LOADING PROCEDURES

Bulk Load Overview

Except for the Regression Terms table, the software lacks the interactive forms that would be required to enter or edit individual table records. Instead, the addition or editing of all data is handled in batches via a bulk-loading mechanism. Other than direct table editing using other software (such as Access itself), this is the only tool available for making changes to the database tables.

All input data (verified as described by Valere 1996) must be in flat text files with fixed-column formats. To decipher these input files the bulk-loading mechanism relies on an external text file named LOADFORM.INI which contains the descriptions of all expected bulk-load formats.

Bulk Load Data Requirements

The external file containing the bulk-load formats must be named LOADFORM.INI and must be located in the same directory as the database itself. This file is formatted like a standard Windows 3.x .INI file, consisting of a section for each bulk-loadable table in the database, such a section beginning with a line containing only the table name enclosed by square brackets. Within a section is a series of lines, one for each loadable field in the table. Each of those lines describes the location and format of the data for that field within a bulk-load input record. Portions of the section relating to the SAMPLES table are reproduced below to illustrate this:

```
[SAMPLES]
FIELD_REFERENCE=CHAR,1,10,10,,1
DATA_SET_ID=CHAR,12,4,15,,1
SAMPLING_ID=CHAR,17,4,20,,
...
START_DATE=DATE,65,6,70,,1
START_TIME=TIME,72,5,76,,
...
SAM_BOTTOM_DEPTH=NUM,98,6,103,2,
SAM_TOP_DEPTH=NUM,105,6,110,2,
DISCRETE_DEPTHS=CHAR,112,80,191,,
OBSERVATION=CHAR,207,240,446,,
```

Here, each line begins with a table field name followed by an equal sign. The text following the equal sign consists of a comma-delimited list of six items providing the location and format information:

1. data type ("CHAR" for text fields, "DATE" or "TIME" for date-time fields and "NUM" for numeric fields)
2. field start position
3. field end position
4. field width (65535 for variable-length text fields)
5. number of decimal places (type "NUM" only)
6. mandatory flag (1 if a required value, null if not)

The complete text of LOADFORM.INI in its present form is provided in Appendix C.

Bulk Load Input File

The software requires that the flat text bulk-load input file the must be named with a .DAT extension. One such file can be used to load data into one and only one database table. Each line in a bulk-load input file must formatted according to the rules provided in LOADFORM.INI for that table, there being one line for each new record to be added to the target table. To prevent accidentally attempting to load the wrong table with a file, it is also required to identify the target table by placing as the very first line in a input file a line like:

```
[TABLE:SAMPLES]
```

where the text between the colon and closing bracket is the name of the target table. Should this line be missing or contain the name of a non-existent table, the target table will default to CODEFLD.

Bulk Load Table Order

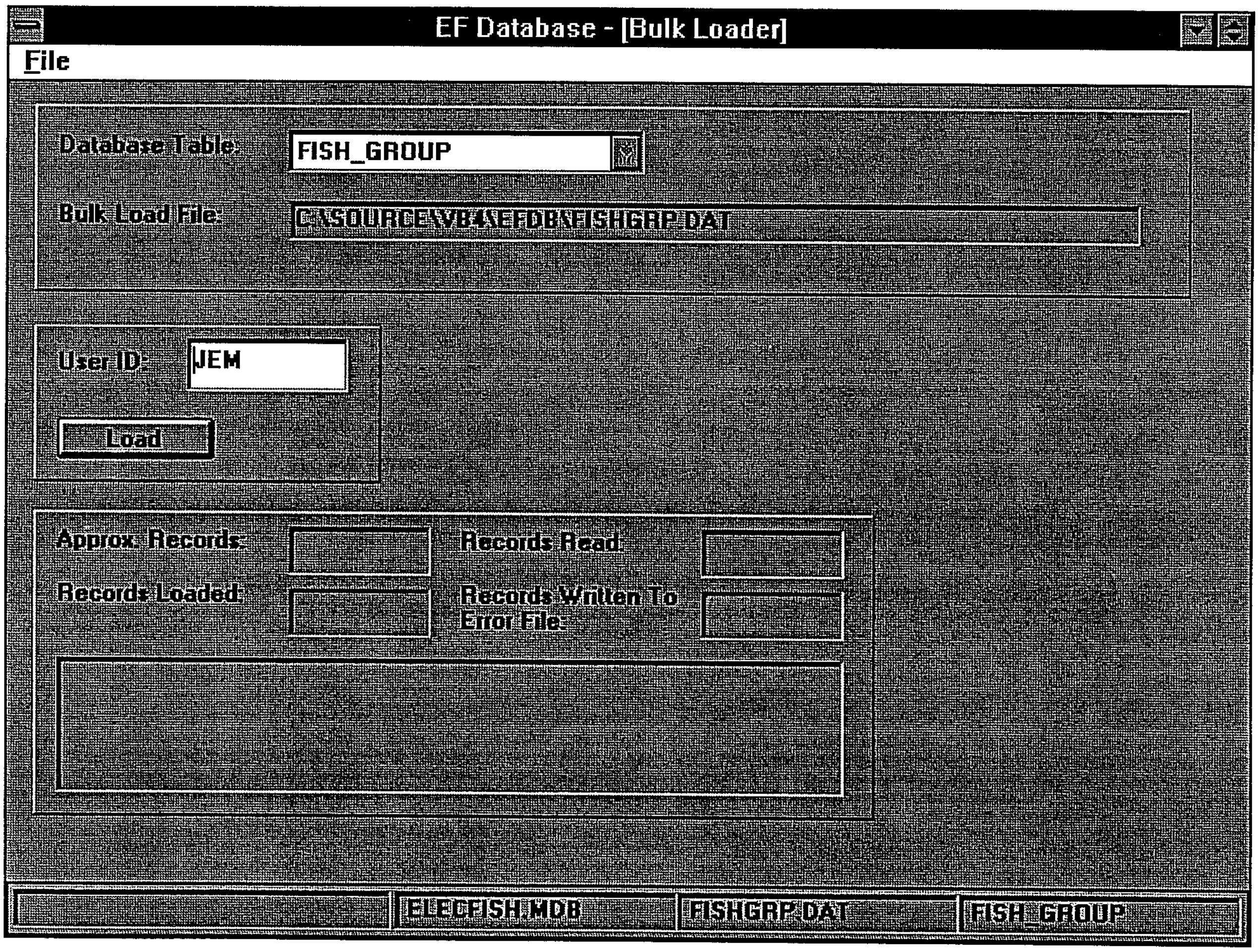
Because of the relational integrity constraints built into the database, the tables must be loaded in order of the table hierarchy depicted in Table 2. Thus if the dataset being loaded contains locations not already present in the database, records for the LOCATION table must be loaded first. If there are new stations, records for the STATION table must then be loaded and so on.

Table 2. Summary of database table loading order.

LOCATION
STATION
SAMPLES
TAXON or ELECTRIC_FISHING or WATER_OBSERVATION
FISH_GROUP
FISH_INDIVIDUAL

In many cases there will be data for only a subset of these tables. The LOCATION, STATION, and (especially) TAXON tables may not require new records in order to load the new data and therefore would not require bulk-load files or consideration in terms of table order. What tables are loaded, however, must still be loaded in the order provided above.

Bulk Load Window



Database Table: FISH_GROUP

Bulk Load File: C:\SOURCE\VB4\EFDB\FISHGRP.DAT

User ID: JEM

ELECFISH.MDB FISHGRP.DAT FISH_GROUP

Figure 2. Appearance of the bulk-load window after selection of an input file.

Initiating a bulk load

To begin a bulk-load select **Bulk Load** from the **File** menu of the main window or click on the **Bulk Load** button. This action will load the bulk-load window where only three actions are required of the user:

1. Select a file containing bulk-load data to be loaded and
2. Enter the user's initials or some other identifier (maximum 10 characters).
3. Click on the **Load** button.

The first step is started automatically when the window loads and can also be started from the **File** menu of the bulk-load window. A standard file dialog box is presented, allowing the user to locate and select any file (in any location) with a .DAT extension. When the file has been selected, its first line will be examined for a target table name, and if a valid name is found, it will appear selected in the combo box at the top of the form.

Once the user's initials have been entered in the **User ID** area of the window, the bulk-load is initiated by clicking on the **Load** button. While the load is going on, a **Stop** button appears which may be clicked on at any time to stop the bulk-load process. Also, as the load is proceeding a status area on the screen indicates errors as they occur and shows:

- the approximate total number of input records
- the number read
- the number successfully loaded
- the number not loaded and written to an error file

Dealing with loading errors

If the number of input records read but not successfully loaded is greater than zero, then two files called ERROR.LOG and ERROR.DAT will have been created in the same directory as the input file. The former contains a brief description (if possible) of the problem associated with each of the input records that failed to load. The latter contains copies of the failed records themselves and can be edited and reloaded. Note that both of these files are recreated for every bulk-load input file that contains bad input records, so if they are to be preserved they must be renamed immediately after the bulk-load event that created them. Also, an attempt to reload data from a file called ERROR.DAT will usually result in an error.

In some cases the error message associated with a bad input line may be something like: "Duplicate primary key". This usually means that the data in question has already been loaded and may not actually represent a problem with the input data itself.

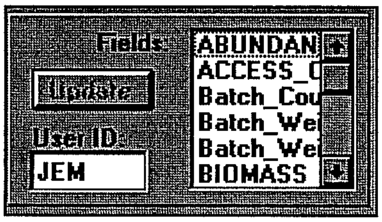
Reversing the effects of a bulk-load session

Every time the bulk-load procedure is run, the system generates an integer number which serves to uniquely identify a bulk-load event. That update identifier is included in every row created in the target table during the bulk-load event, and is stored in the Updates table along with the user's initials, date and time, and other information. Because the Updates table has relationships based on the Update identifier with every loadable table in the database, and because those relationships are enforced with cascading deletes, it is possible to delete a block of rows loaded during a bulk-load event by simply deleting the row with the corresponding update identifier in the Updates table. Be aware, though, that the target table may also have relationships with other tables, so that the effects of deleting a row from the Updates table may cascade beyond the target table into all of its subordinate tables.

Bulk Update Window

It is possible to use the same bulk-loading approach to doing batch updates of a database table. To do this, the software provides a separate bulk-update window which can be activated by clicking on the **Bulk Update** button on the main window. Once activated, the bulk-update window looks and behaves very much like the bulk-load window except that the software allows the user to update a sub-set of the target table's fields for updating.

Adjacent to the **User ID** entry box is a list box:



User ID: JEM

Figure 3. The User ID area of the bulk-update window.

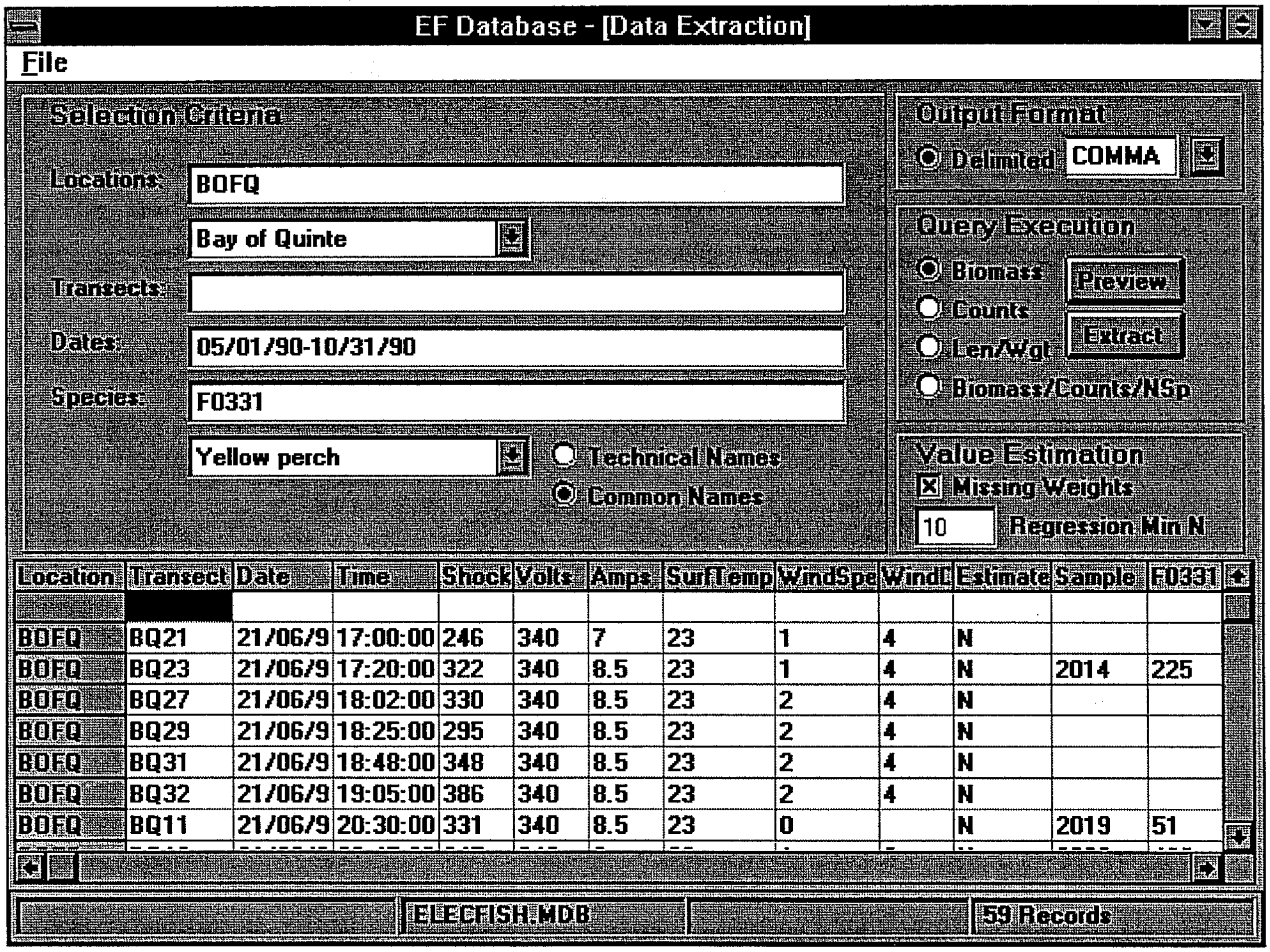
The box will contain a list of all the fields in the table which is the target for the batch update. A field may be selected for updating by clicking on the field's name. More than one field may be selected by clicking on field names while pressing the control or shift keys.

As soon as one or more fields have been selected, the **Update** button becomes active and clicking on that button will initiate the batch update process. Everything that happens from that point is identical to what happens in the bulk-load window.

Note that the software will not allow any attempts to update fields which are key fields in the target table, and will not allow update of other special fields such as the UpdateID or the Field_Reference field in the SAMPLES table.

RETRIEVAL FEATURES

Data Extraction Window



Locations: BOFQ

Bay of Quinte

Transects:

Dates: 05/01/90-10/31/90

Species: F0331

Yellow perch

○ Technical Names

● Common Names

● Delimited COMMA

● Biomass

○ Counts

○ Len/Wgt

○ Biomass/Counts/NSp

☒ Missing Weights

10 Regression Min N

Location	Transect	Date	Time	Shock	Volts	Amps	SurfTemp	WindSpe	WindD	Estimate	Sample	F0331
BOFQ	BQ21	21/06/9	17:00:00	246	340	7	23	1	4	N		
BOFQ	BQ23	21/06/9	17:20:00	322	340	8.5	23	1	4	N	2014	225
BOFQ	BQ27	21/06/9	18:02:00	330	340	8.5	23	2	4	N		
BOFQ	BQ29	21/06/9	18:25:00	295	340	8.5	23	2	4	N		
BOFQ	BQ31	21/06/9	18:48:00	348	340	8.5	23	2	4	N		
BOFQ	BQ32	21/06/9	19:05:00	386	340	8.5	23	2	4	N		
BOFQ	BQ11	21/06/9	20:30:00	331	340	8.5	23	0		N	2019	51

ELECFISH.MDB

59 Records

Figure 4. Appearance of the data extraction window with query parameters and preview results.

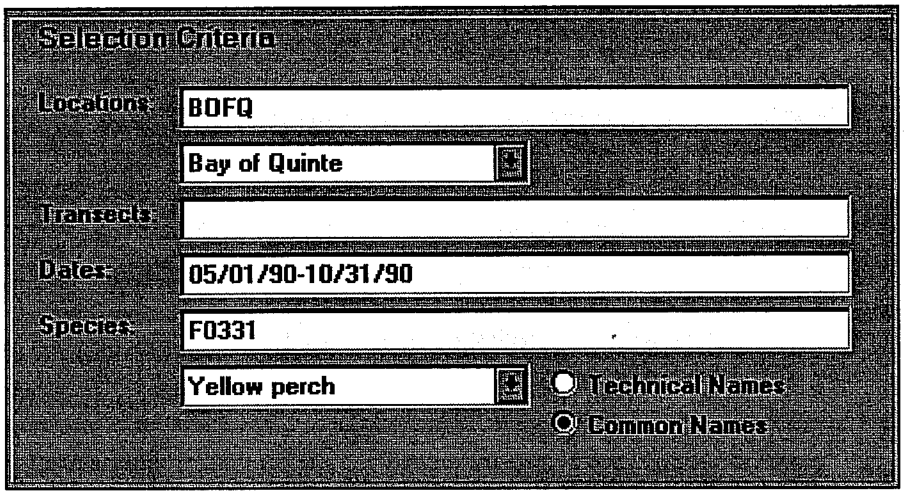
The software provides a means of extracting data of several types from the database and saving the extracted data to external operating system files. This facility is provided by the data-extraction window which is opened from the main window by clicking on the **Extraction** button.

Once the data-extraction window is opened, the user's task is to:

1. construct the parameters of a query which the software will use to retrieve data
2. choose an output format and delimiter
3. choose an extraction query type and
4. set parameters related to estimation of missing weights.

Setting Data Extraction Query Parameters

Query parameters are provided by entering values into one or more of the text boxes available on the **Selection Criteria** portion of the extraction window:



Selection Criteria

Locations: BOFQ

Bay of Quinte

Transects:

Dates: 05/01/90-10/31/90

Species: F0331

Yellow perch

○ Technical Names

● Common Names

Figure 5. Selection Criteria area of the data extraction window.

These boxes are labelled **Locations**, **Transects**, **Dates**, and **Species** and are related to the database as follows:

Table 3. Relationship between extraction query text boxes and the underlying database tables.

Text Box:	Related DB Table:	DB Table Field:
LOCATIONS	LOCATION	LOCATION_ID
TRANSECTS	STATION	STATION_ID
DATES	SAMPLES	START_DATE
SPECIES	TAXON	ACCESS_CODE

If nothing is entered in any of these four boxes, a query will by default retrieve all records from the database. To restrict the records retrieved, for example, to one or more locations, the corresponding location identifiers can be entered in the **Locations** text box. All four text boxes work in the same way, in that criteria may be entered as:

- a single value: value1
- a single range of values: value1 - value2
- a comma-delimited list of values and/or ranges

The query logic combines these things by doing logical ORs between items in any one box and logical ANDs between text boxes.

Some assistance is provided in list construction by the **Locations** and **Species** drop-down lists which provide lists of the location and species codes currently in the database. The **Species** drop-down list may be viewed with either common or technical names for the species.

Query Type

Four different types of queries may be performed: Biomass, Counts, Len/Wgt, and Biomass/Count/NSP. At any one time, one of these may be chosen by clicking on the corresponding radio button in the **Query Execution** area of the extraction window:

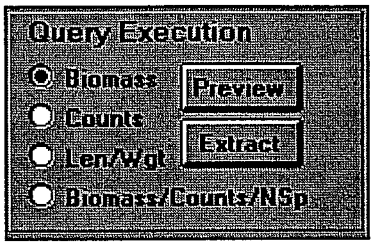


Figure 6. Query Execution area of the data extraction window.

When an extraction is done, this will result in one of four different extraction results:

Biomass extraction result

The output consists of fish biomass by species and transect. Each biomass value represents for one species the sum of all the individual and batch weights obtained in one sampling (electrofishing) effort on a transect. In addition to the species and transect identifiers, each record in the output contains all other information related to the sample. Note that the first line contains labels for the fields in the succeeding lines. The field labelled "Estimate" contains either "N" or "B" to indicate whether the biomass value involved no estimation ("N") or estimation ("B"). The last field or fields are labelled with fish species codes and it is in those fields that the biomass value(s) will be found. An example for just one species follows:

```
Location,Transect,Date,Time,Shock,Volts,Amps,SurfTemp,WindSpeed,WindDir,Estimate,Sample,F0331
BOFQ,BQ21,21/06/90,17:00:00,246,340,7,23,1,4,N,,
BOFQ,BQ23,21/06/90,17:20:00,322,340,8.5,23,1,4,B,2014,225
BOFQ,BQ27,21/06/90,18:02:00,330,340,8.5,23,2,4,N,,
BOFQ,BQ29,21/06/90,18:25:00,295,340,8.5,23,2,4,N,,
BOFQ,BQ31,21/06/90,18:48:00,348,340,8.5,23,2,4,N,,
BOFQ,BQ32,21/06/90,19:05:00,386,340,8.5,23,2,4,N,,
BOFQ,BQ11,21/06/90,20:30:00,331,340,8.5,23,0,,N,2019,51
BOFQ,BQ13,21/06/90,20:45:00,345,340,9,23,1,2,N,2020,409
```

Counts extraction result

The output consists of fish numbers by species and transect. Each fish numbers value represents for one species the sum of all the individual and batch counts obtained in one sampling (electrofishing) effort on a transect. In addition to the species and transect identifiers, each record in the output contains all other information related to the sample. Note that the first line contains labels for the fields in the succeeding lines. The field labelled "Estimate" contains either "N" or "B" to indicate whether the biomass value involved no estimation ("N") or estimation ("B"), neither of which actually means anything in this context. The last field or fields are labelled with fish species codes and it is in those fields that the numbers value(s) will be found. An example for just one species follows:

```
Location,Transect,Date,Time,Shock,Volts,Amps,SurfTemp,WindSpeed,WindDir,Estimate,Sample,F0331
BOFQ,BQ21,21/06/90,17:00:00,246,340,7,23,1,4,N,,
BOFQ,BQ23,21/06/90,17:20:00,322,340,8.5,23,1,4,B,2014,14
BOFQ,BQ27,21/06/90,18:02:00,330,340,8.5,23,2,4,N,,
BOFQ,BQ29,21/06/90,18:25:00,295,340,8.5,23,2,4,N,,
BOFQ,BQ31,21/06/90,18:48:00,348,340,8.5,23,2,4,N,,
BOFQ,BQ32,21/06/90,19:05:00,386,340,8.5,23,2,4,N,,
BOFQ,BQ11,21/06/90,20:30:00,331,340,8.5,23,0,,N,2019,1
BOFQ,BQ13,21/06/90,20:45:00,345,340,9,23,1,2,N,2020,10
```

Len/Wgt extraction result

Each line in the output contains lengths and weights for an individual fish. These are accompanied by identifying information and the count and biomass values for the group of which the individual is part. Note that the first line contains labels for the fields in the succeeding lines. The field labelled "Estimate" contains either "N", "B", or "W" to indicate whether the biomass value involved no estimation ("N"), estimation ("B"), or whether the individual weight was estimated ("W"). An example follows:

```
Sample_ID,Location,Transect,Date,Time,Species,Count,Estimate,Biomass,Number Individuals,Length,Weight
2014,BOFQ,BQ23,21/06/90,17:20:00,F0331,14,B,225,14,83,9
2014,BOFQ,BQ23,21/06/90,17:20:00,F0331,14,B,225,14,132,30
2014,BOFQ,BQ23,21/06/90,17:20:00,F0331,14,B,225,14,78,8
2014,BOFQ,BQ23,21/06/90,17:20:00,F0331,14,W,225,14,116,24
2014,BOFQ,BQ23,21/06/90,17:20:00,F0331,14,B,225,14,76,6
2014,BOFQ,BQ23,21/06/90,17:20:00,F0331,14,B,225,14,72,4
2014,BOFQ,BQ23,21/06/90,17:20:00,F0331,14,B,225,14,71,5
2014,BOFQ,BQ23,21/06/90,17:20:00,F0331,14,B,225,14,138,38
```

Biomass/Count/NSP extraction result

The output consists of total fish biomass, total fish numbers, and species richness by transect. Each total biomass value represents the sum over all species of the individual and batch weights obtained in one sampling (electrofishing) effort on a transect. The total fish numbers value is the corresponding sum of fish counts. The species richness value is the number of species involved in the totals. In addition to the species and transect identifiers, each record in the output contains all other information related to the sample. Note that the first line contains labels for the fields in the succeeding lines. The field labelled "Estimate" contains either "N" or "B" to indicate whether the biomass value involved no estimation ("N") or estimation ("B"). An example follows:

```
Sample,Location,Transect,Date,Time,Shock,Amps,Volts,SurfTemp,Secchi,WindSpeed,WindDir,Estimate,Biomass,
Numbers,NSp
2013,BOFQ,BQ21,21/06/90,17:00:00,246,7,340,23,2.1,1,4,N,11498,5,2
2014,BOFQ,BQ23,21/06/90,17:20:00,322,8.5,340,23,1.25,1,4,B,951,21,4
2015,BOFQ,BQ27,21/06/90,18:02:00,330,8.5,340,23,2,2,4,N,14594,133,4
2016,BOFQ,BQ29,21/06/90,18:25:00,295,8.5,340,23,2.2,2,4,N,144,1,1
2017,BOFQ,BQ31,21/06/90,18:48:00,348,8.5,340,23,,2,4,N,0,0,0
2018,BOFQ,BQ32,21/06/90,19:05:00,386,8.5,340,23,,2,4,N,0,0,0
2019,BOFQ,BQ11,21/06/90,20:30:00,331,8.5,340,23,,0,,N,51,1,1
2020,BOFQ,BQ13,21/06/90,20:45:00,345,9,340,23,,1,2,N,1897,20,4
```

Previewing Query Results and Saving Query Results to a File

Once the query criteria have been entered and the query type has been chosen, the corresponding query results may be viewed by clicking on the **Preview** button in the **Query Execution** area of the extraction window. The query will execute and the results will be returned to a table that will become visible at the bottom of the window. Up to 2000 rows of any one query result set can be viewed in this way.

To save the result set to a file instead, one has to click on the **Extract** button located in the same area of the window. A standard file dialog box will appear in which an output file name with a .EXT extension is selected. When a file has been chosen, the query will execute and the results will be saved to the output file as a delimited text file. The choice of a delimiter is made in the **Output Format** area of the extraction window where there is a drop-down list of the possible delimiter characters: COMMA (default), TAB, and SPACE.

Note that it is not necessary to preview a result set before saving it to a file. The two operations are completely independent.

Missing Value Estimation

The **Value Estimation** area of the extraction window provides a means of controlling the process of providing estimates for missing weights using length-weight regressions. By default if an individual or batch weight is missing from the database, the corresponding biomass values will be set to –1 in a query result set, indicating they are invalid. The software has the ability to provide estimates of missing weights using corresponding lengths in the case of individuals or mean weights in the case of batches.

To have the software do this estimation during a preview or extraction query, the **Missing Weights** checkbox in the **Value Estimation** area must be set to checked if not already so. Beneath that checkbox is a text box labelled **Regression Min N** in which a number (which defaults to 10) may be entered. This determines the minimum sample size a regression must have to be used in the estimation process. If the available data for the location in question will not yield this sample size, the software will pool regression terms across locations in a manner controlled by the **Pooling Tree Tab** of the regression manager window until the sample size is reached or exceeded.

The computation of a mean weight is necessary when a batch weight is missing, in which case it will be estimated as the product of the batch count and the estimated mean weight. Also, the data for a few individual fish are missing both lengths and weights, and an estimated mean weight will be used for such individuals. The software attempts to compute an average weight of the species in question using the individuals in the sample from which it came. If the number of such individuals is less than the **Regression Min N** threshold, the software will use a geometric mean derived from the same pooled regression terms it would otherwise use to estimate a weight from a length.

Finally, weights estimated from a log-log regression of weights and lengths tend to be overestimates, i.e., they are biased. A correction factor can be computed from the same regression term set used to compute the regression (Sprugel 1983). The software computes this correction factor and multiplies all weights estimated from lengths by that factor. Typically this factor is in the range 1.0 to 1.1.

Production of Estimation Reports

After a data extraction has completed, whether weight estimation was asked for or not, an **Extraction Report** item becomes available on the extraction window's **File** menu, providing the means to generate a text file with a .RPT extension containing a summary of what was done for the extraction. This time-stamped report always contains the query details from the extraction window and the type of extraction requested. If weight estimation had been selected as an option, the software will have maintained a history of what was done during the estimation process. That history is reported as the last two sections of the extraction report, e.g.:

```
15/01/97 18:42:37
Extraction report for the following conditions:

Extraction Type: Biomass


Query Details:
Locations> HAMH,BOFQ,KESW
Transects>
Dates> 01/01/90-12/31/95
Species>


Estimation problems: Weight estimation failed for one or more records.
Yield:  1325 RECORDS


Weight estimation: ON
Minimum L-W regression N: 10




Length-Weight regressions employed:

Location Species Uses     N   ICept  Slope     r2 MinL MaxL MinW    MaxW GeoMeanW CorrFact

HAMH     F0186   HAMH  1090 -10.016 2.8738 0.877  128  855   53  12700   2659.7   1.0195
HAMH     F0251   HAMH    29  -0.654 1.1309 0.361   53 1020   90   3260    769.2   1.3973
BOFQ     F0061   BOFQ   143 -10.284 2.7853 0.974   32  150    1     41     11.8   1.0196
KESW     F0194   WORL   394 -11.622 3.0706 0.908   36  164    1     62      4.2   1.0508
KESW     F0196   WORL  1760 -10.411 2.7200 0.746   35  154    1     29      3.8   1.0418
.
.
.

Weights estimated from averages:

Type  Sample Location  Transect     Date Species Uses Weight

INDIV 1902   HAMH          HH18 23/05/90  F0251 REGR  769.2
INDIV 1908   HAMH          HH30 24/05/90  F0251 REGR  769.2
INDIV 2560   HAMH          HH44 07/06/95  F0301 REGR   65.8
INDIV 2560   HAMH          HH44 07/06/95  F0313 MEAN   22.0
.
.
.
BATCH 2629   HAMH          HH44 04/08/95  F0317 MEAN   12.3
BATCH 2715   HAMH          HH41 14/05/92  F0196 MEAN    5.9
BATCH 2717   HAMH          HH43 14/05/92  F0196 REGR    4.4
```

The first three columns of the first section following the query details indicates whether a regression was required for a species-location combination and what pooling location was used. Such a line will be in the report if one or more individuals and or batches required weight estimation for that species and location. The remaining columns give the regression details: sample size, intercept, slope, r-square, minimum and maximum lengths and weights, geometric mean weight, and the correction factor applied to the estimates.

The last section of the report details instances in which weights could not be estimated: either for individuals because the lengths were also missing or for batches because the batch weights were missing. The first column of the report indicates whether the weight was estimated for an individual or a batch and the last two columns give the weight used and whether it was based on the mean of weights for the species in the sample of which it was a part ("MEAN") or whether, because the sample size was too small, the mean used was the geometric mean ("REGR") for the pooling location giving a suitable sample size. The rest of the columns give details about the actual individual or batch involved.

Note that this example shows one other thing which is that for some species there aren't enough individuals in the whole database to do a regression. This is reported as a weight estimation failure on the status line of the extraction window and in the report. The affected species-location combination will appear in the first section of the report following the query details and can be recognized by the absence of all information other than the species and location identifiers.

REGRESSION MANAGER WINDOW

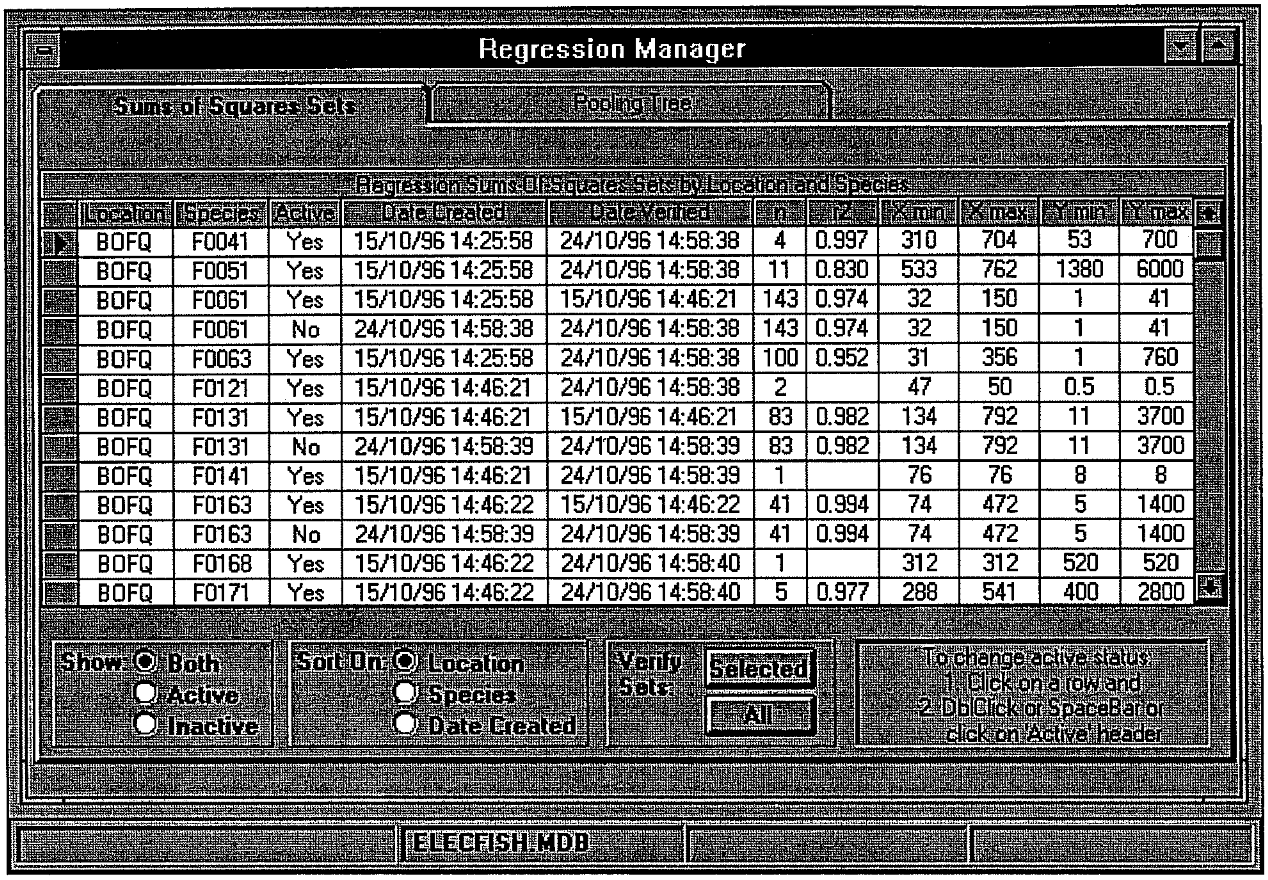
Overview

A field sample may contain individual fish for which no reliable weight could be obtained. This invalidates the total weight for the fish group to which that individual belongs and therefore the biomass for the corresponding sample transect. A substantial portion of the data may be so affected, rendering the extraction query results involving biomass difficult to use. A remedy for this problem is provided by an option of estimate missing individual weights from lengths, using a linear regression of log(weight) on log(length). The regression manager window allows the user to manage a collection of these length-weight regressions.

Each regression in the collection is done by pooling all individuals of a particular species ever collected at one sample location, so that there is one regression for each species-location combination. Occasionally this results in a sample size which is too small to be useful, so that in addition to generating the regressions, the software provides a means for pooling regressions over locations to obtain a desired sample size, and it provides a means for controlling the logic of that pooling process so that only related locations are pooled. The elements of the estimation process are as follows:

1. a minimum regression sample size; set in the **Value Estimation** area of the data extraction window and defaulting to 10
2. sets of regression terms by species and location generated; controlled from the **Sums of Squares Sets Tab** in the regression manager window
3. a pooling location tree defining the relationship between locations and order in which pooling will be done; controlled from the **Pooling Tree Tab** in the regression manager window

Sums of Squares Sets Tab



Regression Sums Of Squares Sets by Location and Species

Location	Species	Active	Date Created	Date Verified	n	r2	X min	X max	Y min	Y max
BOFQ	F0041	Yes	15/10/96 14:25:58	24/10/96 14:58:38	4	0.997	310	704	53	700
BOFQ	F0051	Yes	15/10/96 14:25:58	24/10/96 14:58:38	11	0.830	533	762	1380	6000
BOFQ	F0061	Yes	15/10/96 14:25:58	15/10/96 14:46:21	143	0.974	32	150	1	41
BOFQ	F0061	No	24/10/96 14:58:38	24/10/96 14:58:38	143	0.974	32	150	1	41
BOFQ	F0063	Yes	15/10/96 14:25:58	24/10/96 14:58:38	100	0.952	31	356	1	760
BOFQ	F0121	Yes	15/10/96 14:46:21	24/10/96 14:58:38	2		47	50	0.5	0.5
BOFQ	F0131	Yes	15/10/96 14:46:21	15/10/96 14:46:21	83	0.982	134	792	11	3700
BOFQ	F0131	No	24/10/96 14:58:39	24/10/96 14:58:39	83	0.982	134	792	11	3700
BOFQ	F0141	Yes	15/10/96 14:46:21	24/10/96 14:58:39	1		76	76	8	8
BOFQ	F0163	Yes	15/10/96 14:46:22	15/10/96 14:46:22	41	0.994	74	472	5	1400
BOFQ	F0163	No	24/10/96 14:58:39	24/10/96 14:58:39	41	0.994	74	472	5	1400
BOFQ	F0168	Yes	15/10/96 14:46:22	24/10/96 14:58:40	1		312	312	520	520
BOFQ	F0171	Yes	15/10/96 14:46:22	24/10/96 14:58:40	5	0.977	288	541	400	2800

Show: ● Both
○ Active
○ Inactive

Sort On: ● Location
○ Species
○ Date Created

Verify Sets:

To change active status:
1. Click on a row and
2. DblClick or SpaceBar or click on Active header

Figure 7. Sums of Squares Sets tab of the regression manager window.

Changing the active status of a regression term set

With the regression manager window open, clicking on the **Sums of Squares Sets** tab will display a grid containing the entire set of regression terms currently available, showing what's available, when it was created and updated. As well, the sample size, sample r-squared, and min and maximum lengths and weights (untransformed) are shown for each regression term set. The purpose of this is to allow choices about what are or are not acceptable regression term sets, for example on the basis of the set's r-square value, and to render the regression term sets active or inactive. Only an active set may be employed in the pooling and estimation process. For any one species-location combination only one set may be active at a time.

Within the grid itself the only thing which may be changed is to toggle a regression set's status from active to inactive or vice versa. This may be done in either of three ways after clicking somewhere in the row of the desired regression set:

1. Click on the **Active** column header
2. Double-click directly in the **Active** column cell of the row
3. Press the space bar.

When a status is changed to inactive, nothing spectacular happens, but when changed to active, the software will change the status and then set any active regression sets for that species-location combination to inactive.

At the bottom of the tab, there are two sets of radio buttons labelled **Show** and **Sort**. The options with the former allows displaying both inactive and active sets, active only, or inactive only. The latter allows the sort order of the displayed sets to be by location, species, or date created. These options can simplify dealing with the data, but one restriction applies: the status of a regression term set can not be changed unless the **Show** option is set to both and the **Sort** option is set to either species or location.

Deleting regression term sets

A regression term set may be deleted with these steps:

1. Click anywhere in the row of the set to be deleted.
2. Click on the record selector on the extreme left of the row.
3. Press the delete key.

A prompt will appear asking for confirmation of the delete, and if the response is affirmative, the set will be immediately deleted. There is no undo, though a new set, using the current data, can always be regenerated through the verify process.

Verifying and adding new regression term sets

At the bottom of the regression manager window is an area labelled **Verify Sets** which allows confirmation of existing regression term sets against the current data and updating of the collection of regression sets whenever new data is added to the database or old data changed.

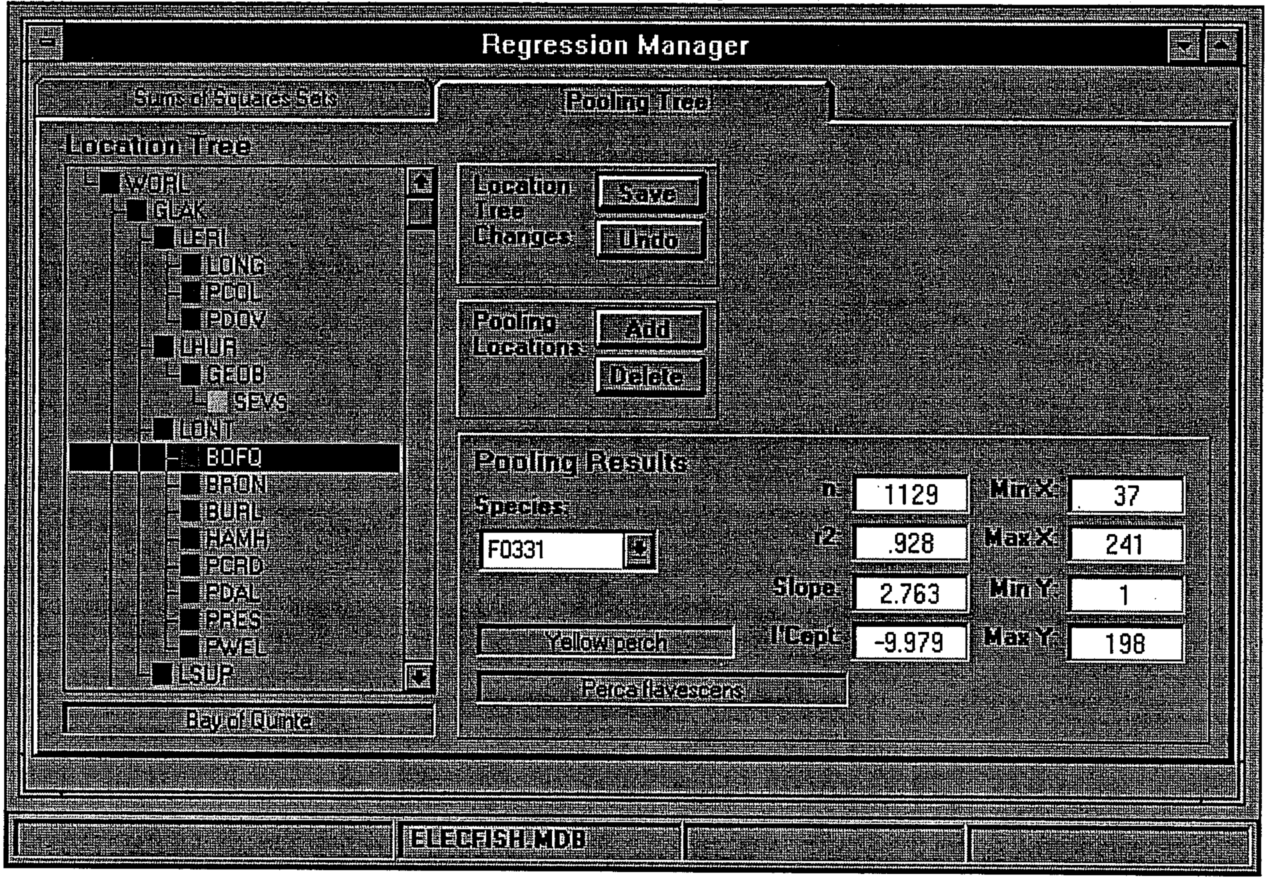
To verify the whole collection, click on the **All** button. This will cause the software to revisit all available data for lengths and weights, recomputing every regression for all species-location combinations. For each species-location combination, the terms in the recomputed regression are compared with those in any corresponding sets of regression terms already in the database. If one stored set matches, the verification date of that set is changed to the current date, but nothing else happens. However, if no matching sets can be found, then the new set is added to the database as a new regression term set with both the creation and verification dates set to the current date and the status as inactive.

A count of such added sets is maintained and reported at the very end of the procedure. The species and location identifiers of the added sets may be determined by sorting the entire collection on creation date, after which the collection may be sorted on either species or location and a determination made as to which set should have an active status. Since by default any new sets are inactive, this verification/update process has no immediate affect on any estimates done during data extraction. An intervention on the part of a user is required to decide which sets should be active and which not.

To restrict this process to only a selected group of species-location combinations, highlight rows in the regression set grid by clicking on the record selectors in those rows. Multiple rows may be selected by doing this while holding the control key down. Click on the **Selected** button, and the same process as described above will take place, but only the highlighted rows will be affected.

Pooling Tree Tab

Clicking on the **Pooling Tree** tab of the regression manager window reveals the part of that window devoted to managing the logic of regression pooling.



Bay of Quinte

Species: F0331

n: 1129

r2: .928

Slope: 2.763

I'Cept: -9.979

Min X: 37

Max X: 241

Min Y: 1

Max Y: 198

Yellow perch

Perca flavescens

Figure 8. Pooling Tree tab of the regression manager window.

When weight estimation is required during an extraction, a set of regression terms is retrieved for the species and location in question and if the sample size for that set meets or exceed the minimum established by the user, that set of terms alone is used to do the regression calculations. If the requirement for a minimum sample size is not met , however, then the software will attempt to pool the regression term set with the set or sets for that species at other locations until the sample size requirement is met or all available data has been used.

At the left of the **Pooling Tree** tab is an outline tree which represents a hierarchical arrangement of location identifiers, usually (though not necessarily) bearing some sort of geographical relationship to one another. Some of those identifiers are accompanied by a green tag, some red, and some yellow. The green locations represent sampling locations actually present in the database for which there may be one or more regression term sets. These locations are necessarily always at the very tips of the tree. The red locations are pooling locations. They have no data associated with them and are created solely for the purpose of grouping green locations in a hierarchical fashion. Finally, the yellow locations are simply red locations which have hidden descendants in the tree.

To pool locations, the software starts at the green location (which has an insufficient sample size) and works it way up the tree one level at a time towards the root, pooling whatever green locations should happen to be descendants of the red location at each level. When a level is reached that produces a pooled sample size meeting the requirements, the pooling stops and the software uses the resulting regression term set to do the regression calculations.

Previewing the regression pooling process

Either by clicking on the text of a location in the pooling tree or by using the arrow keys to move up and down through the tree, a location identifier is selected (highlighted) in the tree. The effect of regression pooling at that location is revealed in the portion of the **Pooling Tree** tab labelled **Pooling Results**. As each location is selected, the corresponding sample size, r-square, and length and weight minima and maxima are displayed for one species which is chosen by selecting a species identifier from the drop-down list in the same area.

Modifying the pooling tree

The pooling tree may be modified by adding new pooling locations, deleting old ones, and by shifting the location identifiers from one place to another within the hierarchy.

A new location may be added by clicking on the **Add** button in the **Pooling Locations** area of the **Pooling Tree** tab. This will cause a small dialog box to appear in which should be entered a location identifier for the new location (one to four characters, e.g., LERI) and a short description of that location (e,g, Lake Erie). Clicking on the **Add** button will then add that location to the tree, if it's not already there. When added, new locations will always appear immediately below the root level in the tree and in alphabetical order among the identifiers at that level.

Locations already in the tree may be deleted by selecting the location and clicking on the **Delete** button in the **Pooling Locations** area. However, a location can only be deleted if: a) it does not have descendants, i.e., it must be at a tip of the tree, and b) if it is a red location, i.e., it does not have data actually associated with it.

Finally, a location identifier may be moved from place to place in the tree by clicking on the text of the location identifier and, while holding the left mouse button down, dragging the identifier and dropping it on the text of another target location identifier. This will cause the location and its descendants to become descendants of that target location. Green locations can be moved, but they can never acquire descendants, i.e., they must always be at the tips of the tree. Also the target location for this drag and drop operation cannot be a descendant of the location being moved.

Saving and undoing changes made to the pooling location tree

The **Pooling Tree** tab has an area labelled **Location Tree Changes**. Whenever one or more changes have been made to the pooling location tree, two buttons labelled **Save** and **Undo** become active. **Undo** will undo the last change made to the tree and can be clicked on repeatedly until all changes have been undone and the displayed tree matches the one currently stored in the database. Save will update the version of the location tree as it's stored in the database and the version displayed will become the currently stored version. At that point both **Save** and **Undo** are no longer possible.

CURRENT DATABASE HOLDINGS

Transect Samples by Locations and Years

Table 4. Summary of current database holdings: Numbers of samples by location and year.

Location	Lake	1988	1989	1990	1992	1993	1994	1995	1996	1997	Total
Long Point	Lake Erie						27				27
Port Colborne	Lake Erie						66				66
Port Dover	Lake Erie						48				48
Bay of Quinte	Lake Ontario		34	59	47						140
Bronte Harbour	Lake Ontario						54				54
Burlington	Lake Ontario						18				18
Hamilton Harbour	Lake Ontario	326		85	90	35		87	160	149	932
Port Credit	Lake Ontario			10							10
Port Dalhousie	Lake Ontario			10			42				52
Port Weller	Lake Ontario						13				13
Prequille Bay	Lake Ontario						30				30
Keswick	Lake Simcoe			8							8
Green Island	Severn Sound							27			27
Hog Bay	Severn Sound			28	6			18			52
Matchedash Bay	Severn Sound			38							38
Penetang Harbour	Severn Sound			84	12						96
Sturgeon Bay	Severn Sound				23						23
Total		326	34	322	178	35	298	132	160	149	1634

Numbers of Records by Table

Table 5. Summary of current database holdings: Numbers of records per table.

Table:	Record Count:
LOCATION	17
STATION	330
SAMPLES	1634
ELECTRIC_FISHING	1634
WATER_OBSERVATION	1634
FISH_GROUP	8292
FISH_INDIVIDUAL	45392

ACKNOWLEDGEMENTS

We thank Mike Stoneman for providing comments and updates on an earlier draft of this report.

REFERENCES

Reference List

MacLeod, W.D., C.K. Minns, A. Mathers and S. Mee. 1995. An evaluation of Biotic Indices and Habitat Suitability Scores for classifying littoral habitats. Can. Manscr. Rep. Fish. Aquat. Sci. 2334: 26 p.

Minns, C.K., V.W. Cairns, R.G. Randall, A. Crowder and A. McLaughlin. 1993. Macrophyte surveys of littoral habitats in Great Lakes' Areas of Concern: The Bay of Quinte, Hamilton Harbour and Severn Sound - 1988-1991. Can. Tech. Rep. Fish. Aquat. Sci. 1936: 60 p.

Minns, C.K., V.W. Cairns, R.G. Randall and J.E. Moore. 1994. An Index of Biotic Integrity (IBI) for fish assemblages in the littoral zone of Great Lakes' Areas of Concern. Can. J. Fish. Aquat. Sci. 51:1804-1822.

Randall, R.G., C.K. Minns, V.W. Cairns, and J.E. Moore. 1993. Effect of habitat degradation on the species composition and biomass of fish in Great Lakes Areas of Concern. Can. Tech. Rep. Fish. Aquat. Sci. 1941:37 p.

Randall, R.G., C.K. Minns, V.W. Cairns, and J.E. Moore. 1996. The relationship between an index of fish production and submerged macrophytes and other habitat features at three littoral areas of the Great Lakes. Can. J. Fish. Aquat. Sci. 53 (Suppl. 1): 35-44.

Sprugel, D.G. 1983. Correcting for bias in log-transformed allometric equations. Ecology 64(1): 209-210.

Valere, B.G. 1996. Productive capacity of littoral hábitats in the Great Lakes: Field sampling procedures (1988-1995). Can. Manusr. Re. Fish. Aquat. Sci. 2384:50 p.

APPENDIX A - Database Table Structures

Data Dictionary

Details with respect to the fields in each of the 11 database tables are provided below. Note that wherever a table or field name is wholly in uppercase, it represents the name of an object as imported from the original GLLFAS Oracle database design by Moneco Inc. Names including lowercase letters were added to that design when the Access version database was created. In all cases, names of key fields appear in boldface type:

CODEFLD

Field Name	Type	Size	Description
ATTRIBUTE	Text	20	Code type ("LIFE_STAGE", "COLLECT_ METHOD", etc.)
CODE	Text	7	Text of code itself
UNITS	Text	10	Units of code (not used)
CODE_DESCRIPTION	Text	60	Detailed description of code
Update_ID	Number (Long)	4	Identifies the bulk load event responsible for a record

ELECTRIC_FISHING

Field Name	Type	Size	Description
SAMPLE_ID	Number (Long)	4	System-generated field sample identifier
SHOCK	Number (Long)	4	Length of shocking time in seconds
AMPS	Number (Double)	8	Mean shocking amperes
VOLTS	Number (Long)	4	Mean shocking voltage
Update_ID	Number (Long)	4	Identifies the bulk load event responsible for a record

FISH_GROUP

Field Name	Type	Size	Description
SAMPLE_ID	Number (Long)	4	System-generated field sample identifier
ACCESS_CODE	Text	5	Species code
COUNT	Number (Long)	4	Total count of individual fish plus the batch count
ABUNDANCE_UNITS	Text	8	Units of the count
BIOMASS	Number (Double)	8	Total weight of individual fish plus the batch weight
BIOMASS_UNITS	Text	8	Units of the biomass value
Batch_Count	Number (Long)	4	The batch count alone
Batch_Weight	Number (Double)	8	The batch weight alone
Batch_Weight_Units	Text	8	Units of the batch weight
Update_ID	Number (Long)	4	Identifies the bulk load event responsible for a record

FISH_INDIVIDUAL

Field Name	Type	Size	Description
SAMPLE_ID	Number (Long)	4	System-generated field sample identifier

ACCESS_CODE	Text	5	Species code
INDIVIDUAL_ID	Text	3	Identifies an individual fish within a group
LIFE_STAGE	Text	3	Code identifying an individual's life stage
WEIGHT	Number (Double)	8	Weight of an individual fish
WEIGHT_UNIT	Text	8	Units of the weight value
LENGTH	Number (Double)	8	Length of an individual fish
LENGTH_UNITS	Text	8	Units of the length value
LENGTH_CODE	Text	3	Code identifying the type of length measurement
AGE	Number (Long)	4	Age of an individual fish in years
AGE_METHOD	Text	3	Code identifying the method used to obtain the age
MATURITY	Text	3	Code identifying the reproductive state of an individual
SEX	Text	3	Code identifying the sex of an individual fish
SEX_METHOD	Text	3	Code identifying the method used to determine the sex
OBSERVATION	Text	240	Miscellaneous descriptive text
Update_ID	Number (Long)	4	Identifies the bulk load event responsible for a record

LOCATION

Field Name	Type	Size	Description
LOCATION_ID	Text	4	Identifies a geographical area containing a number of stations
Parent_Location_ID	Text	4	Location ID of this parent's location in the regression tree
LOCATION_NAME	Text	30	Name of the location
GAZETEER_NAME	Text	30	Official NTS name of the location
LATITUDE_DEGREE	Number (Long)	4	Latitude degrees for a reference point
LATITUDE_MINUTE	Number (Long)	4	Latitude minutes for a reference point
LATITUDE_SECOND	Number (Double)	8	Latitude seconds for a reference point
LONGITUDE_DEGREE	Number (Long)	4	Longitude degrees for a reference point
LONGITUDE_MINUTE	Number (Long)	4	Longitude minutes for a reference point
LONGITUDE_SECOND	Number (Double)	8	Longitude seconds for a reference point
DESCRIPTION	Memo	-	Detailed description of the location
Update_ID	Number (Long)	4	Identifies the bulk load event responsible for a record

SAMPLES

Field Name	Type	Size	Description
SAMPLE_ID	Number (Long)	4	System-generated field sample identifier (a counter)
LOCATION_ID	Text	4	Geographical area in which the sample was collected
STATION_ID	Text	5	Point at which the sample was collected
SAMPLE_TYPE	Text	5	Identifies the sample type (always "FISH")
COLLECT_METHOD	Text	5	Code identifying the sample collection method
PROCESS_METHOD	Text	5	Code identifying the sample processing method (not used)
FIELD_REFERENCE	Text	10	Sample identifier used at the time of field collection
START_DATE	Date/Time	8	Date sample collection began
START_TIME	Date/Time	8	Time sample collection began
END_DATE	Date/Time	8	Date sample collection ended (not used)
END_TIME	Date/Time	8	Time sample collection ended (not used)
SAMP_DURATION	Date/Time	8	Time required to collect sample (not used)
OBSERVATION	Text	240	Miscellaneous descriptive text
Update_ID	Number (Long)	4	Identifies the bulk load event responsible for a record

STATION

Field Name	Type	Size	Description
LOCATION_ID	Text	4	Identifies a geographical area containing a number of stations
STATION_ID	Text	5	Identifies a point at which a sample was collected
STATION_TYPE	Text	3	Code identifying the type of use this station has
STATION_NAME	Text	30	Name of the station
LATITUDE_DEGREE	Number (Long)	4	Latitude degrees for a reference point
LATITUDE_MINUTE	Number (Long)	4	Latitude minutes for a reference point
LATITUDE_SECOND	Number (Double)	8	Latitude seconds for a reference point
LONGITUDE_DEGREE	Number (Long)	4	Longitude degrees for a reference point
LONGITUDE_MINUTE	Number (Long)	4	Longitude minutes for a reference point
LONGITUDE_SECOND	Number (Double)	8	Longitude seconds for a reference point
GEOREFERENCE_ACCURAC	Text	3	Code identifying the method used to obtain the position of the reference point
DESCRIPTION	Memo	-	Detailed description of the station
Update_ID	Number (Long)	4	Identifies the bulk load event responsible for a record

TAXON

Field Name	Type	Size	Description
ACCESS_CODE	Text	5	Species code
GENUS	Text	20	Genus name
SPECIES	Text	20	Species name
COMMON_NAME	Text	30	Common name
PHYLUM	Text	20	Phylum name
CLASS	Text	20	Class name
ORDER	Text	20	Order name
SUB_ORDER	Text	20	Sub-order name
FAMILY	Text	20	Family name
SUB_FAMILY	Text	20	Sub-family name
TRIBE	Text	20	Tribe name
SUB_GENUS	Text	20	Sub-genus name
CF_GROUP	Text	20	Referral group name
UPDATE_DATE	Date/Time	8	(not used)
ACCOUNT	Text	30	(not used)
VERSION	Number (Long)	4	(not used)
SPECIES_TYPE	Text	4	always "FISH"
Update_ID	Number (Long)	4	Identifies the bulk load event responsible for a record

Updates

Field Name	Type	Size	Description
Update_ID	Number (Long)	4	Identifies the bulk load event responsible for a group of records
User_ID	Text	10	Identifies the person responsible for loading a group of records
Date	Text	9	Date a group of records was loaded
Time	Text	5	Time a group of records was loaded
Description	Memo	-	Rationale for loading a group of records
Table_ID	Text	30	Name of the table into which the records were loaded
Record_Count	Number (Long)	4	Count of records successfully loaded

WATER_OBSERVATION

Field Name	Type	Size	Description
SAMPLE_ID	Number (Long)	4	System-generated field sample identifier
SECCHI_BW_DEPTH	Number (Double)	8	Black-and-white Secchi depth in metres
WIND_DIRECTION	Text	3	Wind direction code (1=N,2=E,3=S,4=W)
WIND_SPEED	Text	3	Wind speed code
SURFACE_TEMP	Number (Double)	8	Surface temperature in degrees C
CONDUCTIVITY	Number (Double)	8	Surface conductivity in umho/cm
OBSERVATION	Text	240	Miscellaneous descriptive text
Update_ID	Number (Long)	4	Identifies the bulk load event responsible for a record

Regression Terms

Field Name	Type	Size	Description
Regression_ID	Number (Long)	4	System-generated identifier identifying a set of length-weight regression terms
LOCATION_ID	Text	4	Geographical area to which the terms apply
ACCESS_CODE	Text	5	Species code of the species to which the terms apply
Active_Flag	Yes/No	1	Indicates whether or not these terms are to be used in L-W regressions
Count	Number (Long)	4	Regression n
SumX	Number (Double)	8	Regression sum of x (logged lengths)
SumY	Number (Double)	8	Regression sum of y (logged weights)
SumX2	Number (Double)	8	Regression sum of x2
SumY2	Number (Double)	8	Regression sum of y2
SumXY	Number (Double)	8	Regression sum of xy
MinX	Number (Single)	4	Minimum length (unlogged)
MaxX	Number (Single)	4	Maximum length
MinY	Number (Single)	4	Minimum weight (unlogged)
MaxY	Number (Single)	4	Maximum weight
R-Square	Number (Double)	8	r2 computed over this set of n points
Create_TimeStamp	Date/Time	8	Time the regression set was computed
Verify_TimeStamp	Date/Time	8	Time the regression set was last verified
Max_Update_ID	Number (Long)	4	Maximum Update_ID in the set of records used for this set

APPENDIX B - Database Code Summary

Database Codes

The following table sets out the complete contents of the CODEFLD table, showing the names and descriptions of all codes currently in use within the database. Note that an "attribute" listed in the first column of this table corresponds to a field name in one or more database tables. Values stored in such a field must be a code and the code should appear in the CODEFLD database table next to the corresponding attribute name. This table is provided for information only. In practice, there is no validation of codes on input.

ATTRIBUTE	CODE	UNITS	CODE_DESCRIPTION	Update_ID
ABUNDANCE_UNITS	NUMBERS		Number of individuals in catch	67
AGE_METHOD	IVD		Invalid age determination	67
COLLECT_METHOD	EF100		Electrofishing on 100 metre transect	67
GEOREFERENCE_ACCURAC	APX		Approximate location	67
GEOREFERENCE_ACCURAC	LOC		Unknown - Assigned coordinates of parent location	68
LENGTH_CODE	FRK		Fork Length	68
LENGTH_CODE	TOT		Total length	67
LENGTH_UNITS	mm		millimetres	67
LIFE_STAGE	ADS		Adult	67
LIFE_STAGE	IMM		Immature	68
LIFE_STAGE	UNK		Unknown	68
SAMPLE_TYPE	FISH		Fish specimens or tissue	67
SEX	F		Female	68
SEX	M		Male	67
SEX	U		Unknown	68
STATION_TYPE	EFX		Electrofishing transect	68
STATION_TYPE	FSH		Fishing station	67
WEIGHT_UNITS	gm		grams	67

APPENDIX C - Listing of Bulk Load Formats

Bulk-Load Formats

Bulk-load formats for all database tables are contained in the file called LOADFORM.INI. That file is broken into sections each of which begins with the name of one of the database tables in square brackets. The rest of the section consists of a series of field descriptions which when taken together describe how every line in an input file must be formatted to successfully load that table. Each field description begins with a table field name followed by an equal sign. The text following the equal sign consists of a comma-delimited list of six items providing the columnar location and format information:

1. data type ("CHAR" for text fields, "DATE" or "TIME" for date-time fields and "NUM" for numeric fields)
2. field start position
3. field end position
4. field width (65535 for variable-length text fields)
5. number of decimal places (type "NUM" only)
6. mandatory flag (1 if a required value, null if not)

The following is a simple list of the complete contents of LOADFORM.INI with blank lines inserted to improve readability. Note that the are references to some tables (DATASET, PROJECT, GLLFAS_MEMBERS and PROJECT_GLLFAS_MEMBERS) not currently implemented in the Access database:

```
[TAXON]
ACCESS_CODE=CHAR,1,5,5,,1
PHYLUM=CHAR,7,20,26,,
CLASS=CHAR,28,20,47,,
ORDER=CHAR,49,20,68,,
SUB_ORDER=CHAR,70,20,89,,
FAMILY=CHAR,91,20,110,,
SUB_FAMILY=CHAR,112,20,131,,
TRIBE=CHAR,133,20,152,,
GENUS=CHAR,154,20,173,,
SUB_GENUS=CHAR,175,20,194,,
SPECIES=CHAR,196,20,215,,
```

```
CF_GROUP=CHAR,217,20,236,,
SPECIES_TYPE=CHAR,238,4,241,,
COMMON_NAME=CHAR,243,30,272,,

[CODEFLD]
ATTRIBUTE=CHAR,1,20,20,,1
CODE=CHAR,22,7,28,,1
UNITS=CHAR,30,10,39,,
CODE_DESCRIPTION=CHAR,41,60,100,,

[DATA_SET]
DATA_SET_ID=CHAR,1,4,4,,1
DATA_SET_NAME=CHAR,6,80,85,,1
DATA_SET_OWNER=CHAR,87,4,90,,1
START_DATE=DATE,92,6,97,,
END_DATE=DATE,99,6,104,,
PROJECT_ID=CHAR,106,4,109,,1
DESCRIPTION=CHAR,111,65535,,,

[ELECTRIC_FISHING]
FIELD_REFERENCE=CHAR,1,10,10,,1
DATA_SET_ID=CHAR,12,4,15,,1
SHOCK=NUM,17,5,21,,
AMPS=NUM,23,4,26,1,
VOLTS=NUM,28,4,31,,

[FISH_GROUP]
FIELD_REFERENCE=CHAR,1,10,10,,1
DATA_SET_ID=CHAR,12,4,15,,1
ACCESS_CODE=CHAR,17,5,21,,1
LIFE_STAGE=CHAR,23,3,25,,1
COUNT=NUM,27,4,30,,
ABUNDANCE_UNITS=CHAR,32,8,39,,
BIOMASS=NUM,41,9,49,2,
BIOMASS_UNITS=CHAR,51,8,58,,
BATCH_WEIGHT=NUM,60,9,68,2,
BATCH_WEIGHT_UNITS=CHAR,70,8,77,,

[FISH_INDIVIDUAL]
FIELD_REFERENCE=CHAR,1,10,10,,1
DATA_SET_ID=CHAR,12,4,15,,1
ACCESS_CODE=CHAR,17,5,21,,1
LIFE_STAGE=CHAR,23,3,25,,1
INDIVIDUAL_ID=CHAR,27,3,29,,1
FISH_COMPOSITE_ID=CHAR,31,3,33,,
CSP_SAMPLE_ID=CHAR,35,6,40,,
WEIGHT=NUM,42,9,50,2,
WEIGHT_UNITS=CHAR,52,8,59,,
```

```
LENGTH=NUM,61,7,67,2,
LENGTH_UNITS=CHAR,69,8,76,,
LENGTH_CODE=CHAR,78,3,80,,
AGE=NUM,82,2,83,,
AGE_METHOD=CHAR,85,3,87,,
MATURITY=CHAR,89,3,91,,
SEX=CHAR,93,3,95,,
SEX_METHOD=CHAR,97,3,99,,
OBSERVATION=CHAR,101,240,340,,

[GLLFAS_MEMBERS]
GLLFAS_MEMBER_ID=CHAR,1,4,4,,1
GLLFAS_MEMBER_NAME=CHAR,6,30,35,,1
POSITION=CHAR,37,30,66,,
START_DATE=DATE,68,6,73,,
END_DATE=DATE,75,6,80,,
DFO_STAFFER=CHAR,82,1,82,,1

[LOCATION]
LOCATION_ID=CHAR,1,4,4,,1
LOCATION_NAME=CHAR,6,30,35,,
GAZETEER_NAME=CHAR,37,30,66,,
LATITUDE_DEGREE=NUM,68,2,69,,1
LATITUDE_MINUTE=NUM,71,2,72,,1
LATITUDE_SECOND=NUM,74,5,78,2,1
LONGITUDE_DEGREE=NUM,80,4,83,,1
LONGITUDE_MINUTE=NUM,85,2,86,,1
LONGITUDE_SECOND=NUM,88,5,92,2,1
DESCRIPTION=CHAR,94,65535,,,

[PROJECT]
PROJECT_ID=CHAR,1,4,4,,1
PROJECT_NAME=CHAR,6,60,65,,
PROJECT_LEADER=CHAR,67,4,70,,
START_DATE=DATE,72,6,77,,
END_DATE=DATE,79,6,84,,
DESCRIPTION=CHAR,86,65535,,,

[PROJECT_GLLFAS_MEMBERS]
GLLFAS_MEMBER_ID=CHAR,1,4,4,,1
PROJECT_ID=CHAR,6,4,9,,1

[SAMPLES]
FIELD_REFERENCE=CHAR,1,10,10,,1
DATA_SET_ID=CHAR,12,4,15,,1
SAMPLING_ID=CHAR,17,4,20,,
COMPOSITE_ID=CHAR,22,3,24,,
PROJECT_ID=CHAR,26,4,29,,
```

```
LOCATION_ID=CHAR,31,4,34,,1
STATION_ID=CHAR,36,5,40,,1
SAMPLE_TYPE=CHAR,42,5,46,,1
COLLECT_METHOD=CHAR,48,5,52,,1
COLLECT_AUTHORITY=CHAR,54,4,57,,
PROCESS_METHOD=CHAR,59,5,63,,
START_DATE=DATE,65,6,70,,1
START_TIME=TIME,72,5,76,,
END_DATE=DATE,78,6,83,,
END_TIME=TIME,85,5,89,,
SAMP_DURATION=TIME,91,6,96,,
SAM_BOTTOM_DEPTH=NUM,98,6,103,2,
SAM_TOP_DEPTH=NUM,105,6,110,2,
DISCRETE_DEPTHS=CHAR,112,80,191,,
NUM_DISCRETE=NUM,193,2,194,,
STRATA=CHAR,196,3,198,,
NUM_REPLICATES=NUM,200,2,201,,
ANALYST_ID=CHAR,203,3,205,,
OBSERVATION=CHAR,207,240,446,,

[STATION]
LOCATION_ID=CHAR,1,4,4,,1
LATITUDE_DEGREE=NUM,6,2,7,,1
LATITUDE_MINUTE=NUM,9,2,10,,1
LATITUDE_SECOND=NUM,12,5,16,2,1
LONGITUDE_DEGREE=NUM,18,4,21,,1
LONGITUDE_MINUTE=NUM,23,2,24,,1
LONGITUDE_SECOND=NUM,26,5,30,2,1
STATION_ID=CHAR,32,5,36,,1
STATION_TYPE=CHAR,38,3,40,,
STATION_NAME=CHAR,42,30,71,,
TRANSECT_ID=CHAR,73,2,74,,
GEOREFERENCE_ACCURAC=CHAR,76,3,78,,
NAQUADAT_STN_#=CHAR,80,5,84,,
DESCRIPTION=CHAR,86,65535,,,

[WATER_OBSERVATION]
FIELD_REFERENCE=CHAR,1,10,10,,1
DATA_SET_ID=CHAR,12,4,15,,1
SECCHI_BW_DEPTH=NUM,17,5,21,2,
SECCHI_WHITE_DEPTH=NUM,23,5,27,2,
WIND_DIRECTION=CHAR,29,3,31,,
WIND_SPEED=CHAR,33,3,35,,
WIND_SPEED_NUM=NUM,37,3,39,,
CLOUD_COVER=NUM,41,2,42,,
WATER_COLOUR_NUM=NUM,44,2,45,,
WATER_COLOUR=CHAR,47,25,71,,
EPILIMNION_DEPTH=NUM,73,6,78,2,
```

```
WAVE_HEIGHT=NUM,80,4,83,2,
N_SQUARED=NUM,85,4,88,,
MEAN_EPI_TEMP=NUM,90,5,94,2,
SURFACE_TEMP=NUM,96,5,100,2,
MEAN_META_ZEU_TEMP=NUM,102,5,106,2,
LIGHT_ATTEN_COEFFICIENT=NUM,108,5,112,2,
TWENTYM_TEMP=NUM,114,5,118,2,
TWENTYFIVEM_TEMP=NUM,120,5,124,2,
MEAN_META_TEMP=NUM,126,5,130,2,
META_BOTTOM_TEMP=NUM,132,5,136,2,
OBSERVATION=CHAR,138,240,377,,
CONDUCTIVITY=NUM,379,6,383,1,
```