EA3892

North and Central Coast Salmon Database and Analysis System User Manual

Prepared for:

Pacific Salmon Foundation

Prepared by:

Wendell Challenger, Tony Mochizuki, Karl English, and Yury Bychkov

LGL Limited
environmental research associates
9768 Second Street
Sidney, BC, V8L 3Y8

September 2018

LGL LIMITED
environmental research associates

NCC Salmon Database and Analysis System User Manual

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

blank page

1. ABOUT THIS GUIDE

1.1 Intended Audience

This manual is intended as a reference for anyone wishing to operate the North and Central Coast (NCC) Salmon Database and Analysis System.

1.2 Assumptions Made

It is assumed that anyone using this manual:

- Is familiar with the on-going North and Central Coast Salmon Database and Analysis project.
- Has a basic understanding of database theory and practice and an intermediate understanding of how to use Microsoft Access 2010 or newer.
- Has a basic understanding of computer programming and the R Computing Environment (R Core Team 2018).
- Has successfully completed the system setup tasks (Appendix A)

1.3 Related Documentation

The following documentation provides more background and details related to the North and Central Coast Salmon Database and Analysis project:

English, K.K., C. Noble and C. Carr-Harris. 2018. Skeena sockeye in-river run reconstruction analysis model and analysis results for 1982-2017. Report for Pacific Salmon Foundation. 45 p.

English, K.K., D. Peacock, W. Challenger, C. Noble, I. Beveridge, D. Robichaud, K. Beach, C. Carr-Harris, S. Davies and E. Hertz. 2018. Review of North and Central Coast salmon indicator streams and estimating escapement, catch and run size for each salmon conservation unit. Report for Pacific Salmon Foundation and Fisheries and Oceans, Canada. 108 p.

1.4 Conventions Used

Several acronyms and abbreviations are used throughout this document and the System's databases. These items can be found in the following table:

Table 1-1. The System's acronym and abbreviation definitions.

Acronym / Abbreviation	Full Word or Phrase
BC	British Columbia
Cdn	Canadian
CO	Coho Salmon
CU	Conservation Unit
CWT	Coded Wire Tag
DFO	Department of Fisheries and Oceans
ER	Exploitation Rate
FOS	Fisheries Operating System
Frm	Form
MS Access	Microsoft Access database software
MS Excel	Microsoft Excel software
NBSRR	Northern Boundary Sockeye Run Reconstruction
NCC	North and Central Coast
NCCC	North Coast and Central Coast
NCCSDB	North and Central Coast Salmon Database (R-package)
NuSEDS	DFO's new Salmon Escapement Database System
PK	Pink Salmon
PSRR	Pink Salmon Run Reconstruction
QA	Quality Assessment
Qry	Query
SEL	Sockeye Salmon (Lake type)
SER	Sockeye Salmon (River type)
SSIR	Skeena Sockeye In-River (model)
SX	Sockeye Salmon
TRTC	Total Return to Canada

1.5 Contacting Us

LGL Limited
9768 Second Street
Sidney, British Columbia
V8L 3Y8 Canada

Phone: 250-656-0127
Fax: 250-655-4761
Web: www.lgl.com

2. OVERVIEW

This chapter describes the North and Central Coast (NCC) Salmon Database and Analysis System and its components, and is intended to familiarize you with the System and its terminology.

2.1 NCC Salmon Database and Analysis System

The NCC Salmon Database System (the System) was initially developed in 2003 to organize all the information needed to compute total escapement, catch, run size and exploitation rates (ERs) for each salmon species for Statistical Areas 01-10 (English et al. 2004a, 2004b, 2006). The System was expanded in 2009 (English et al. 2009) to produce similar estimates for each of the newly defined NCC Conservation Units (CUs) of Pacific salmon (Holtby and Ciruna 2007). Further refinements to the NCC Salmon Database and related analytical models for 1980-2010 were made from 2011-2012 (English et al. 2013a, 2013b). In 2012-2013, the time series for Area 04 salmon CUs were extended back to 1950s or 1960s and combined with age composition data to provide the estimates by brood year needed for stock-recruit analysis (English 2013; Korman and English 2013). In 2018, all MS Access programming logic was audited and migrated to the R computing environment. Overall, the System is based on a modular design, making it easy to operate and maintain as well as providing the flexibility to expand the system when needed.

Within this system there are seven major components (Figure 2-1):

1. Escapement and stream data from the DFO NuSEDS database and associated meta data;
2. Catch numbers from the DFO FOS database;
3. Age composition data from DFO databases;
4. Exploitation rates derived from various external models and coded wire tag analyses (see English et al. 2018 descriptions and references related to these analyses);
5. Select TRTC results from external analyses;
6. Analysis parameter values; and
7. The R programs required to combine all these components into annual estimates of escapement, catch and run size for each CU are described in Appendix B.

The database is designed to facilitate the uploading of key information components when they are provided in standardized data formats. Data insertions and subsequent computations are computed within the R computing environment (R Core Team 2018) using routines (i.e., functions) from the custom NCCSDB package that was developed to accompany the NCC Salmon Database (Figure 2-1 and Table 2-1).

The NCC Salmon Database is contained in a single MS Access file (*NCC Salmon Database v2.accdb*) which is updated via the main update script (*nccsdb-analysis.R*) and with all the data import settings contained in an associated settings script (*import-settings.R*; Table 2-1). A single MS Access file (*NCC Salmon Database v2.accdb*) is produced containing all the raw data and analytical components of the System. An in-depth review on how to run setup and run R analysis script, along with documentation on the supporting NCCSDB R-package can be found in Appendix A, Appendix B, and Appendix C respectively.

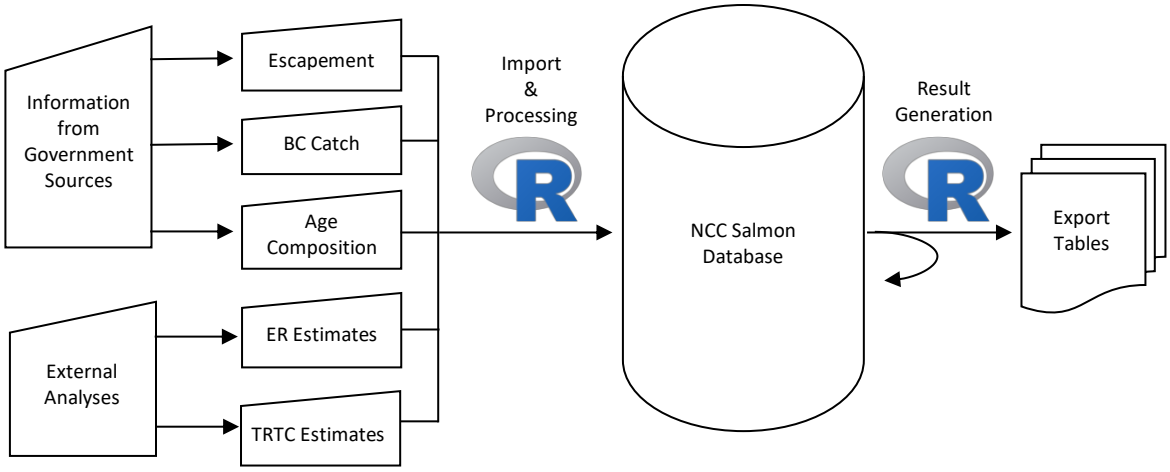

Figure 2-1. High-level data flow diagram of the North and Central Coast Salmon Database.

Table 2-1. Main files required to update the NCC Salmon Database.

File/Directory	Type	Description/Notes
NCC Salmon Database v2.accdb	MS Access Database file	Microsoft Access database.
/data	File directory	File directory containing all the original source data files referred to by *import-settings.R.*
import-settings.R	Text File	R script containing all the file paths to data sources and associated import settings, which is read by the main analysis script.
nccsdb-analysis.R	Text File	Main analysis script which contains the main analysis parameter settings (e.g., reconstruction years, legacy support), loads the data into the database, runs the results, then s exports the results to file as well as saving results with the database itself.
/NCCSDB	File directory	Custom R-package directory containing all the R routines required to import data into the database, run analyses and export results. See Appendix C for further details.

3. IMPORTING PRIMARY DATA SOURCES

Primary data sources and estimates from external analyses are imported into the NCC Salmon Database through an R script and associated R routines found in the NCCSDB R package (Figure 2-1). Once all required data has been inserted into the NCC Salmon Database various processing functions are called to compile, copy and otherwise process the imported data into the data tables used from the final output calculations.

Data sources and destination tables associated with updating the primary data sources and associated R routines are indicated in Figure 3-1 with further details about the source files and the affected database tables are outlined in Table 3-1.

Primary data has also been broken down into four distinct classes, escapement and stream meta data (see Section 3.1), FOS catch data (3.3), age composition data (see Section 3.3), and data associated with the Pink & Chum external analysis model, which was recently converted to R (see Section 3.4).

It should also be noted that the System uses a number of internally derived fields to represent species designations, unique stream by species combinations and conservation units (Table 3-2). The species designation used within the System (i.e., SpeciesId) distinguishes between odd and even year Pink, but do not distinguish between river and lake type Sockeye (Table 3-3). The IndexId field provides a unique stream by species identifier that distinguishes between even and odd year Pink populations, unlike the NuSEDS POP_ID field and finally the internally derived CU field provides a unique conservation unit identifier for each species (Table 3-2).

These conversions are implemented during the various data import routines included in the NCCSDB package (Appendix B and Appendix C).

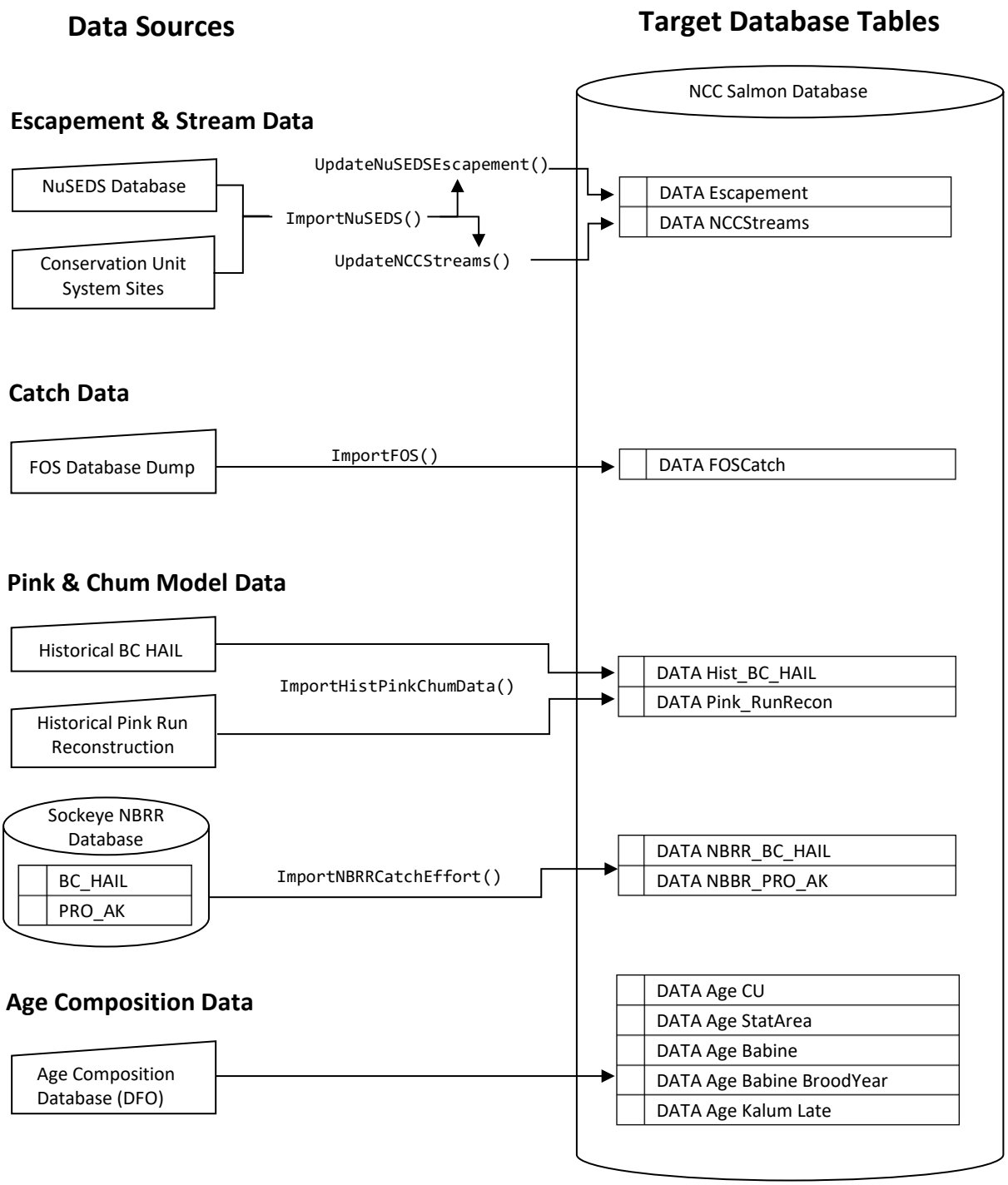

Figure 3-1. Relationships between primary data sources, their respective System database tables, and associated R routines used to import or update primary data sources.

Table 3-1. Summary of the data tables associated with primary data sources.

Report Section	Table Name	Table Schema	Overview	Data Source	Further Details
3.1 Stream and Escapement Data	DATA Escapement	Table **3-4**	NuSEDS stream escapement for the NCC	DFO NuSEDS database (April 4, 2018 release) see for further details	Appendix D
	DATA NCCStreams	Table **3-5**	Master list of NCC streams	*Conservation_Unit_System_Sites.csv* provided with the NuSEDS data release (April 4, 2018 release)	Appendix E
3.2 Catch Data	DATA FOSCatch	Table **3-6**	DFO FOS catch data	CatchData_1954-2017_Areas1-10.xlsx (July 27, 2018 release)	Appendix F
3.3 Age Composition Data	DATA Age CU	Table **3-7**	Conservation unit age composition	DFO age composition data *2016-age-composition-data.xlsx*	Appendix G
	DATA Age StatArea	Table **3-8**	Statistical area age composition		
	DATA Age Babine	Table **3-9**	Babine Sockeye age composition data		
	DATA Age Babine BroodYear	Table **3-10**	Babine Sockeye age composition by brood year. Included for reference only		
	DATA Age Kalum Late	Table **3-11**	Kalum Chinook Late-run age composition included for reference only		
3.4 Pink and Chum Model Data	DATA Hist_BC_HAIL	Table **3-12**	Historical BC HAIL data 1954-1982	*Pink&ChumModel_1954-2014_8Dec2015a_CN.edits.xlsm* (Dave Peacock, per comm.)	
	DATA NBRR_BC_HAIL	Table **3-13**	Current BC HAIL data	Northern Boundary Sockeye Run Reconstruction Database (NBSBR) (Alexander et al. 2010)	
	DATA NBRR_PRO_AK	Table **3-14**	Current Alaska effort and catch		
	DATA Pink_RunRecon	Table **3-15**	Pink run reconstruction results 1982-1995	*Pink&ChumModel_1954-2014_8Dec2015a_CN.edits.xlsm* (Gazey and English, 2000)	

Table 3-2. Internal fields derived from primary data.

Field	Purpose	Further Details
SpeciesId	Species identification code used within the NCC Salmon Database.	Table 3-3
IndexId	Provides a unique stream by species identifier used within the System instead of the NuSEDS POP_ID field in order to distinguish between even and odd year Pink populations. While the NuSEDS POP_ID field was designed to identify specific population of salmon by species, spawning stream and run-timing, no distinction is made between even and odd year Pink populations.	Section 3.1
CU	Provides a unique conservation unit identifier and is a combination of the System SpeciesId field and the CU_INDEX field found in the Conservation Unit System Sites file distributed with NuSEDS data.	Section 5.4
StatArea	Indicates the statistical area which follows the NuSEDS statistical areas, but collapses some of the additional breakdowns provided by NuSEDS (i.e., 3A and 3B are designated as 03, and 4A, 4B, 4C, and 4D are designated as 04). All areas are also given a leading zero if only one digit is used to make all StatArea codes two digits. As such, most results use the following statistical area breakdown: 01, 02E, 02W, 03, 04, 05, 06, 07, 08, 09, 10. The only exception are the Chinook TRTC results, which are derived from an external analysis (see Section 4.2) and provide a further breakdown of Area 09 into 9S and 9W.	Section 5.4

Table 3-3. Species code designation associations between the NCC Salmon Database, NuSEDS and the Conservation Unit System Sites file distributed with NuSEDS.

Species	NCC Salmon Database	NuSEDS	Conservation Unit System Sites
Chinook	CN	Chinook	CK
Chum	CM	Chum	CM
Coho	CO	Coho	CO
Pink (Even Year)	PKe	Pink	PKE
Pink (Odd Year)	PKo		PKO
Sockeye (Lake Type)	SX	Sockeye	SEL
Sockeye (River Type)			SER

3.1 Stream and Escapement Data[1]

All the escapement data for BC salmon streams was obtained from the DFO NuSEDS database and provided in a standardized format (see Appendix D). The key data fields extracted from the NuSEDS data dumps are listed in [DATA Escapement] and [DATA NCCStreams] tables (Table 3-4 and Table 3-5 respectively). The [DATA Escapement] field contains escapement data that was derived from NuSEDS (full details are provided in Appendix D) and [DATA NCCStreams] contains

[1] The source data formats and procedures for importing the escapement, catch and age composition data from DFO databases are described in Appendix D.

the complete list of streams in the NCC derived from the Conservation Unit System Sites file distributed with NuSEDS (see Appendix E for full details).

The list of indicator streams has also been maintained in the "NCC Streams File" (see Section 5.1). Periodically, this file has been used by DFO to update an "indicator stream" field in the NuSEDS database. The procedure for approving and conducting regular updates to the indicator stream field in the NuSEDS database needs to be established. Until then, the list of indicator streams used in the NCC Salmon Database will be the latest version of the "NCC Streams File" file as approved by the senior North Coast stock assessment biologist.

Table 3-4. [DATA Escapement] Annual NuSEDS escapement estimates for NCC streams.

Field Name	Data Type	Used in Analysis	Description
Id	Counter	No	Unique NuSEDS record identifier
SpeciesId	Text	Yes	Species identification code
IndexId	Text	Yes	Unique stream by species identifier (combination of SpeciesID and PopId fields)
PopId	Integer	No	NuSEDS POP_ID field. Unique number used to identify a specific population of salmon defined by species, spawning stream and run-timing
Year	Integer	Yes	Calendar year
StatArea	Text	No	Statistical area
Returns	Integer	Yes	Number of salmon returning to the stream (i.e., escapement)

Table 3-5. [DATA NCCStreams] Master list of NCC streams along with associated stream meta data derived from the Conservation Unit System Sites file.

Field Name	Data Type	Used in Analysis	Description
SpeciesId	Text	Yes	Species identification code
IndexId	Text	Yes	Unique stream by species identifier (combination of SpeciesId and PopId fields). Used internally instead of PopId in order to deal with Pink even and odd year populations that have the same PopId
PopId	Integer	No	NuSEDS POP_ID field. Unique number used to identify a specific population of salmon defined by species, spawning stream and run-timing
Records	Integer	No	Total Number of NuSEDS records available.
Surveys	Integer	No	Number of NuSEDS records with
Active	Logical	Yes	Indicates whether the stream has non-zero escapement records
SYSTEM_SITE	Text	No	The name of the waterbody. Originally from NUSEDS but not necessarily the same. Name priority was BC gazette > BC provincial alias > DFO alias1 > DFO alias2. Same as Waterbody Name above
StatArea	Text	Yes	Statistical area

Field Name	Data Type	Used in Analysis	Description
CU	Text	Yes	Conservation unit code
CU_NAME	Text	Yes	The assigned name of the Conservation Unit. Note that this name does not identify the species.
CU_ACRO	Text	No	The acronym assigned to the CU
CU_LATITUDE	Number	No	Centroid of the CU
CU_LONGITUDE	Number	No	Centroid of the CU
SPECIES_QUALIFIED	Text	Yes	Conservation Unit acronym used to describe the species of salmon for which the escapement estimate is for, e.g.: CK - Chinook Salmon CM - Chum Salmon CO - Coho Salmon PKE - Even Year Pink Salmon PKO - Odd Year Pink Salmon SEL - Lake Type Sockeye Salmon SER - River or Ocean Type Sockeye Salmon
YLAT	Number	No	Location of the mouth of the waterbody if flowing, or the centroid if not.
XLONG	Number	No	Location of the mouth of the waterbody if flowing, or the centroid if not.
FAZ_ACRO	Text	No	Acronym of the freshwater adaptive zone
MAZ_ACRO	Text	No	Acronym of the marine adaptive zone
JAZ_ACRO	Text	No	Acronym of the joint adaptive zone
SITE_ID	Integer	No	A unique numeric code identifying the site. One site can have numerous populations
GFE_ID	Integer	No	Numeric code identifying the waterbody. From NUSEDS with some additions and modifications. Same as Stream_Id listed above
NUMBER_OF_SITES	Integer	No	The number of POP_ID's in the CU. The value of this field and the latitude and longitude fields following come from a pivot table on the "spp_site_pivots" worksheets and are the total counts (#sites) and averages of lat and long for the sites in the CU. If you make changes to the sites in any CU refresh the pivot table to update the values of these fields.
CU_TYPE	Text	No	There are currently six CU types, i.e.,Current, Bin, VREQ[Bin], VREQ[Current], VREQ[Extirpated] and Extirpated based upon Blair Holtby's Rev 4.0 Conservation Unit data refresh.
SBJ_ID	Integer	No	NuSEDS code: 1=Sockeye, 2=Coho, 3=Pink, 4=Chum, 5=Chinook
IS_INDICATOR	Text	No	Is "Y" if this POP_ID has been identified by Area experts as an "indicator" population. This field may be out of date relative the indicators used in [LOOKUP IndicatorStreams].
OL_GRP_NM	Text	No	A name assigned to the SACC Outlook Group
OL_GRP_N	Integer	No	A number assigned to the SACC Outlook Group.
AREA	Text	No	Area Description
ISENH	Text	No	Is "Y" if either SEP_ENH or NUSEDS_Enh is "Y". If NUSEDS_Enh is "UNK" and SEP_ENH is "N" then is "UNK". Otherwise is "N"

Field Name	Data Type	Used in Analysis	Description
COMMENTS	Text	No	Any comments. This field was not added until late Sept 2012 and contains comments pertaining to changes required by 10Sep2012 snapshot only. And, as of Apr 2013, comments related to rev.4 sites changes/additions/etc.
GFE_ID_IN_NUSEDS	Text	No	Is "Y" if the GFE_ID appears in the Geo_Features table of NUSEDS
POP_ID_IN_NUSEDS	Text	No	Is "Y" if the POP_ID appears in NUSEDS. Note that there should be a reference somewhere to the NUSEDS snapshot against which the verification was done.
CMNT	Text	No	Comment Describing the version of the CU data
EFFECTIVE_DT	Text	No	Date that the Conservation unit data became effective in NuSEDS

3.2 Catch Data

All the commercial fishery harvest data for Sockeye, Pink and Chum salmon was obtained from the DFO FOS database. These data are used to estimate ER only for statistical areas where most of the catch of a specific species is comprised of fish originating from that statistical area (e.g., Area 08 Chum). For Sockeye, Pink and Chum salmon stocks caught in mixed-stock fisheries or several statistical areas, ER rates were derived from external run reconstruction analyses. Exploitation rate estimates for Chinook and Coho salmon were derived from external analyses of coded wire tag recovery data.

A single MS Access table contains all the salmon catch data used in the NCC Salmon Database. Table 3-6 provides the format for the MS Access catch table. These catch data were obtained from the FOS database using the standardized format provided in Appendix F.

Table 3-6. [DATA FOSCatch] FOS Catch data by statistical area, species, gear and statistical week.

Field Name	Data Type	Used in Analysis	Description
RawId	Integer	No	FOS unique record identifier
SourceId	Integer	No	Source of data (not used in analysis)
FisheryId	Integer	No	Type of fishery (not used in analysis)
Year	Integer	Yes	Calendar year
StatWeek	Integer	Yes	Statistical week
StatArea	text	Yes	Statistical area
SubAreaId	text	No	Sub area
GearId	text	Yes	Gear type identification code (i.e., GN = Gillnet, IF = Food, Social, and Ceremonial Fishery, SN = Seine Net, TR = Trawling, UN = Unknown)
SpeciesId	text	Yes	Species identification code
N	Integer	Yes	Number of salmon caught and kept

3.3 Age Composition Data

Age composition data is needed for each statistical area by species and each CU in order to produce the estimates of total return by age and brood year for stock-recruit analysis. Estimates of the average annual age composition for each salmon species returning to each statistical area and CU were derived from the Pacific Region Salmon Age Dataset (Brian Spilsted, pers. comm.). Additional data on the annual age composition of Sockeye salmon returns to the Nass and Skeena watershed were provided by Richard Alexander and Steve Cox-Rogers, respectively. For those CU with minimal or no age composition estimates, surrogate CU or statistical area estimates were identified for these CUs (Dave Peacock, pers. comm.). The formats for the age composition tables in the NCC Salmon Database (DATA Age StatArea, DATA Age CU, DATA Babine Age ReturnYear, DATA Babine Age BroodYear, DATA Kalum Late Age) are provided in Table 3-7, Table 3-8, Table 3-9, Table 3-10, and Table 3-11 respectively. Appendix G provides the format of the raw data received from DFO.

Table 3-7. [DATA Age CU] Age composition data by species and conservation unit.

Field Name	Data Type	Used in Analysis	Description
ID	Integer	No	Unique record identifier
Species	Text	Yes	Species identification code
Sp-CU_Name	Text	No	Species and conservation unit
CU	Text	Yes	Conservation unit code
HasAgeData	Text	Yes	Flag indicating whether age data is available, used for filtering.
Age 2	Number	Yes	Proportion of species specific CU returns that are of age 2
Age 3	Number	Yes	Proportion of species specific CU returns that are of age 3
Age 4	Number	Yes	Proportion of species specific CU returns that are of age 4
Age 5	Number	Yes	Proportion of species specific CU returns that are of age 5
Age 6	Number	Yes	Proportion of species specific CU returns that are of age 6
Age 7	Number	Yes	Proportion of species specific CU returns that are of age 7
CUstoUse	Text	No	Conservation units to use
CU2	Text	Yes	Substitute age composition data from a different conservation unit or statistical area

Table 3-8. [DATA Age StatArea] Age composition data by species and statistical area.

Field Name	Data Type	Used in Analysis	Description
ID	Integer	No	Unique record identifier
StatArea	Text	Yes	Statistical area
SPECIES	Text	Yes	Species identification code
SumOfYear1	Number	No	N/A
SumOfYear2	Number	No	N/A
SumOfYear3	Number	No	N/A
SumOfYear4	Number	No	N/A

SumOfYear5	Number	No	N/A
SumOfYear6	Number	No	N/A
SumOfYear7	Number	No	N/A
SumOfYear8	Number	No	N/A
Age 2	Number	Yes	Proportion of statistical area returns that are age 2
Age 3	Number	Yes	Proportion of statistical area returns that are age 3
Age 4	Number	Yes	Proportion of statistical area returns that are age 4
Age 5	Number	Yes	Proportion of statistical area returns that are age 5
Age 6	Number	Yes	Proportion of statistical area returns that are age 6
Age 7	Number	Yes	Proportion of statistical area returns that are age 7

Table 3-9. [DATA Age Babine] Age composition data by year for Babine Sockeye.

Field Name	Data Type	Used in Analysis	Description
Year	Number	No	Calendar year
Age 3	Number	No	Proportion of stock returns that are age 2
Age 4	Number	No	Proportion of stock returns that are age 3
Age 5	Number	No	Proportion of stock returns that are age 4

Table 3-10. [DATA Age Babine BroodYear] Age composition data by year for Babine Sockeye.

Field Name	Data Type	Used in Analysis	Description
BroodYear	Integer	No	Brood Year
Age 3	Number	No	Proportion of stock returns that are age 2
Age 4	Number	No	Proportion of stock returns that are age 3
Age 5	Number	No	Proportion of stock returns that are age 4

Table 3-11. [DATA Age Kalum Late] Age composition data by year for Kalum Late-timing Chinook.

Field Name	Data Type	Used in Analysis	Description
FISCAL_YEAR	Integer	No	Calendar year
Statistical Area	Text	No	Statistical area
CU Name	Text	No	Conservation unit name
CU	Text	No	Conservation unit code
SPECIES	Text	No	Species identification code
Year1	Number	No	Number of returns of age 1
Year2	Number	No	Number of returns of age 2
Year3	Number	No	Number of returns of age 3
Year4	Number	No	Number of returns of age 4
Year5	Number	No	Number of returns of age 5
Year6	Number	No	Number of returns of age 6
Year7	Number	No	Number of returns of age 7

Year8	Number	No	Number of returns of age 8
Total	Number	No	Total number of returns
Yr2	Number	No	Percentage of returns of age 1
Yr3	Number	No	Percentage of returns of age 2
Yr4	Number	No	Percentage of returns of age 3
Yr5	Number	No	Percentage of returns of age 4
Yr6	Number	No	Percentage of returns of age 5
Yr7	Number	No	Percentage of returns of age 6
Comment	Text	No	Comments

3.4 Pink and Chum Model Data

The Pink & Chum model estimates exploitation rate for Pink and Chum salmon in Area 3, Area 4 and Area 5 (Challenger et al. 2018). The model uses a combination of historical and current catch and effort data from the BC Hail database, harvest and exploitation rates estimated from the 1982-1995 Pink Salmon Run Reconstruction (PSRR) analysis (Gazey and English 2000) and Chum Salmon harvest rates determined from the Chum Model (English et al. 2018). With the recent conversion of the model to R all associated data is now kept in data tables within the System. Chum Model output is stored as part of the external analysis data (see Section 4.1.2).

3.4.1 Historical BC Hail Data

The publicly available BC Hail database contains weekly effort and catch values for 1982 onwards only. As such, BC catch and effort data prior to 1982 were provided by Dave Peacock and are contained in the [DATA Hist_BC_Hail] data table (Table 3-12). The data provides total weekly fishing effort (i.e., boat-days) by gear type (i.e., gillnet or seine only).

Table 3-12 [DATA Hist_BC_HAIL] Historical BC Hail data (1954-1981).

Field Name	Data Type	Used in Analysis	Description
StatArea	Text	Yes	Statistical area
Gear	Text	Yes	Fishing gear type (i.e., gillnet or seine)
AKWeek	Integer	Yes	Alaska week
Year	Integer	Yes	Calendar year
BoatDays	Integer	Yes	Measure of fishing effort in boat-days

3.4.2 Sockeye NBRR Catch and Effort Tables

The current BC Hail database contains catch and total fishing effort (boat-days) for 1982 to 2017 and was extracted from Northern Boundary Sockeye Run Reconstruction (NBSRR) database and stored in the [DATA NBRR_BC_HAIL] table (Table 3-13).

Table 3-13 [DATA NBRR_BC_HAIL] Current BC Hail effort data (i.e., 1982-present).

Field Name	Data Type	Used in Analysis	Description
ID	Integer	No	Unique record identifier

Field Name	Data Type	Used in Analysis	Description
YEAR	Integer	Yes	Calendar year
AREA	Integer	Yes	Statistical Area (i.e., 1 to 10 without the leading zero typically used in the NCC Salmon Database)
Hail_Area	Text	No	Description of the area surveyed
FisheryNo	Integer	Yes	
Fishery	Text	No	Fishery name
RealDate	Text	No	Text string indicating survey date
AK_WEEK	Integer	Yes	Alaska week
GEAR	Text	Yes	Fishing gear type identification code (i.e., 01 = gillnet, 02 = seine, 03 = trawl) only gear types 01 and 02 are used by the Pink & Chum model.
BOATS	Integer	Yes	Number of boats
SOCKEYE	Integer	No	Sockeye catch kept
SOCKEYE_RELD	Integer	No	Sockeye catch released
COHO	Integer	No	Coho catch kept
COHO_RELD	Integer	No	Coho catch released
PINK	Integer	Yes	Pink catch kept - only pink catch is used in the Pink & Chum model.
PINK_RELD	Integer	No	Pink catch released
CHUM	Integer	No	Chum catch kept
CHUM_RELD	Integer	No	Chum catch released
CHINOOK	Integer	No	Chinook catch kept
CHINOOK_RELD	Integer	No	Chinook catch released
JACKS	Integer	No	Chinook jacks caught
STHD	Integer	No	Steelhead caught

Fishing effort data for Southern Southeast Alaska purse seine fisheries was also obtained from the database and was stored in the [DATA NBRR_PRO_AK] table (Table 3-14).

Table 3-14 [DATA NBRR_PRO_AK] Current Alaska effort and catch data (i.e., 1982-present).

Field Name	Data Type	Used in Analysis	Description
recno	Integer	No	Unique record identifier
Year	Integer	Yes	Calendar year
FisheryNo	Integer	Yes	Fishing district, only districts 20, 21, 26, 27 are used in the Pink & Chum model
Fishery	Text	No	Fishery name
AKwk	Integer	Yes	Alaska week
OpeningDate	Text	No	Opening date of the fishery (format yyyy-mm-dd)
ClosingDate	Text	No	Closing date of the fishery (format yyyy-mm-dd)
OpenDays	Integer	No	Number of days the fishery was open

Gear	Text	Yes	Fishing gear type code (i.e., 01 = gillnet, 02 = seine, 03 = trawl) only gear types 01 and 02 are used by the Pink & Chum model.
Boats	Integer	Yes	Number of boats
Sockeye	Integer	No	Sockeye catch kept
Coho	Integer	No	Coho catch kept
Pink	Integer	No	Pink catch kept
Chum	Integer	No	Chum catch kept
Chinook	Integer	No	Chinook catch kept

3.4.3 Pink Run Reconstruction Estimates

Pink reconstruction results (Gazey and English 2000) is stored in the [DATA Pink_RunRecon] data table and covers 1982-1995 and includes effort data (i.e., boat-days) along with harvest and exploitation rate estimates for the Area 03 and Area 04 Pink salmon (Table 3-15).

Table 3-15 [DATA Pink_RunRecon] Estimates from the Pink salmon run reconstruction data (1982-1995).

Field Name	Data Type	Used in Analysis	Description
StatArea	Text	Yes	Statistical area
Year	Integer	Yes	Calendar year
Effort	Integer	Yes	Fishing effort in boat days
AreaHR	Number	Yes	Statistical area specific harvest rate
CanadianHR	Number	Yes	Canadian harvest rate
AlaskaHR	Number	Yes	Alaska harvest rateAlaskan exploitation rate.
CanadianER	Number	No	Canadian exploitation rate
TotalER	Number	No	Total exploitation rate

4. IMPORTING EXTERNAL ANALYSIS RESULTS AND DATA PROCESSING

Analysis output (see Section 6) is a combination of output results calculated from the primary data sources (see Section 3) and results derived from analyses external to the System (i.e., the current section). English et al. (2018) provides a full description of each. External analyses are used to provide select estimates of exploitation rates (see Section 4.1) as well as complete TRTC output in some instances (see Section 4.2). These data are individually stored in INPUT tables (Table 4-1), which are later combined prior to computing output results (see Section 4.3). Data files or source associated with each external analysis result is also indicated in Table 4-1. Further details outlining how the data is imported into the System is covered in Appendix B.

Table 4-1. List of input tables used by the NCC Salmon Database System and their source.

Report Section	Table Name	Details	Notes	Source
4.1 Externally Estimated Exploitation Rates	INPUT SX Nass Cdn ER	Table **4-2**	Nass Sockeye Canadian exploitation rates	Sockeye Exploitation Rate Summary1982-2017_20July2018.xlsm
	INPUT SX Nass Total ER	Table **4-3**	Nass Sockeye total exploitation rates	
	INPUT SX Skeena Cdn ER	Table **4-4**	Skeena Sockeye Canadian exploitation rates	
	INPUT SX Skeena Total ER	Table **4-5**	Skeena Sockeye total exploitation rates	
	INPUT Chum HRs	Table **4-6**	Chum harvest rates for Area 03-05	Chum macro v10_20180622.xlsm
	INPUT Pink ERs	Table **4-7**	Pink exploitation rates model Area 03-05	Pink & Chum Model
	INPUT Chum ERs	Table **4-8**	Chum exploitation rates Area 03-05	
	INPUT Coho CUs+ERs Cdn	Table **4-9**	Coho Canadian exploitation rates	Coho_CUs+ERs+Babine_1954-2017_KKE.xlsx
	INPUT Coho CUs+ERs Total	Table **4-10**	Coho total exploitation rates	
4.2 External Estimated TRTC Results	INPUT SockeyeSkeenaNassReconstruction	Table **4-11**	Skeena (Area 03) and Nass (Area 04) Sockeye external TRTC estimates	Sockeye Exploitation Rate Summary1982-2017_20July2018.xlsm
	INPUT Sockeye Babine CUs	Table **4-11**	Babine CU Sockeye external TRTC estimates	
	INPUT Coho Babine CU TRTC	Table **4-11**	Babine CU Coho external TRTC estimates	Coho_CUs+ERs+Babine_1954-2017_KKE.xlsx
	INPUT Chinook TRTC	Table **4-11**	Chinook TRTC estimates for statistical areas 03-04, 06, 08, 9S, 9W, and 10	Chinook TRTC NCC_31July2018.xlsx
	INPUT Chinook CU TRTC	Table **4-11**	Chinook TRTC estimates for CUs CN_36 - CN_40, CN_42, CN_57 - CN_58	

4.1 Externally Estimated Exploitation Rates

Many of the ERs used in the NCC Salmon Database to derive catch estimates for the various CUs originate from analyses external to the NCC Salmon Database. Sockeye salmon ERs for all Nass and Skeena Sockeye salmon CUs were derived from the Northern Boundary Sockeye Run Reconstruction Model (Gazey and English 2000; English et al. 2004a; Alexander et al. 2010) and the Skeena Sockeye In-River model (English et al. 2013b). The external analysis procedures used to derive the ERs for Area 03-05 Pink and Chum salmon CUs and all NCC Chinook and Coho salmon CUs are described in English et al. (2013a). English (2013) describes the methods used to extend the ER time series for Area 04 (Skeena River) salmon CUs back to 1960 for Sockeye and to 1954 for Pink, Chum and Coho salmon.

Exploitation rate data is first imported into the respective tables using a general import routine and import settings detailing which Excel workbook, worksheet and cells to read the data from (see Appendix B for further details). Each data source also has its own custom routine to cleanup and prepare the data prior to insertion (Figure 4-1 provides an overview of all R routines used). After all external exploitation rates and TRTC results have been imported, post processing routines are called to reconfigure and combine estimates prior to computing the final output (see Section 4.3)

For each of the species, a separate table contains the link between each conservation unit and the best available set of annual exploitation rate estimates for that conservation unit (see Section 5.2).

For Sockeye salmon, four tables (Table 4-2– Table 4-5) are used to provide the Canadian and Total ERs for each statistical area, group of statistical areas or stock group in Areas 03 and 04.

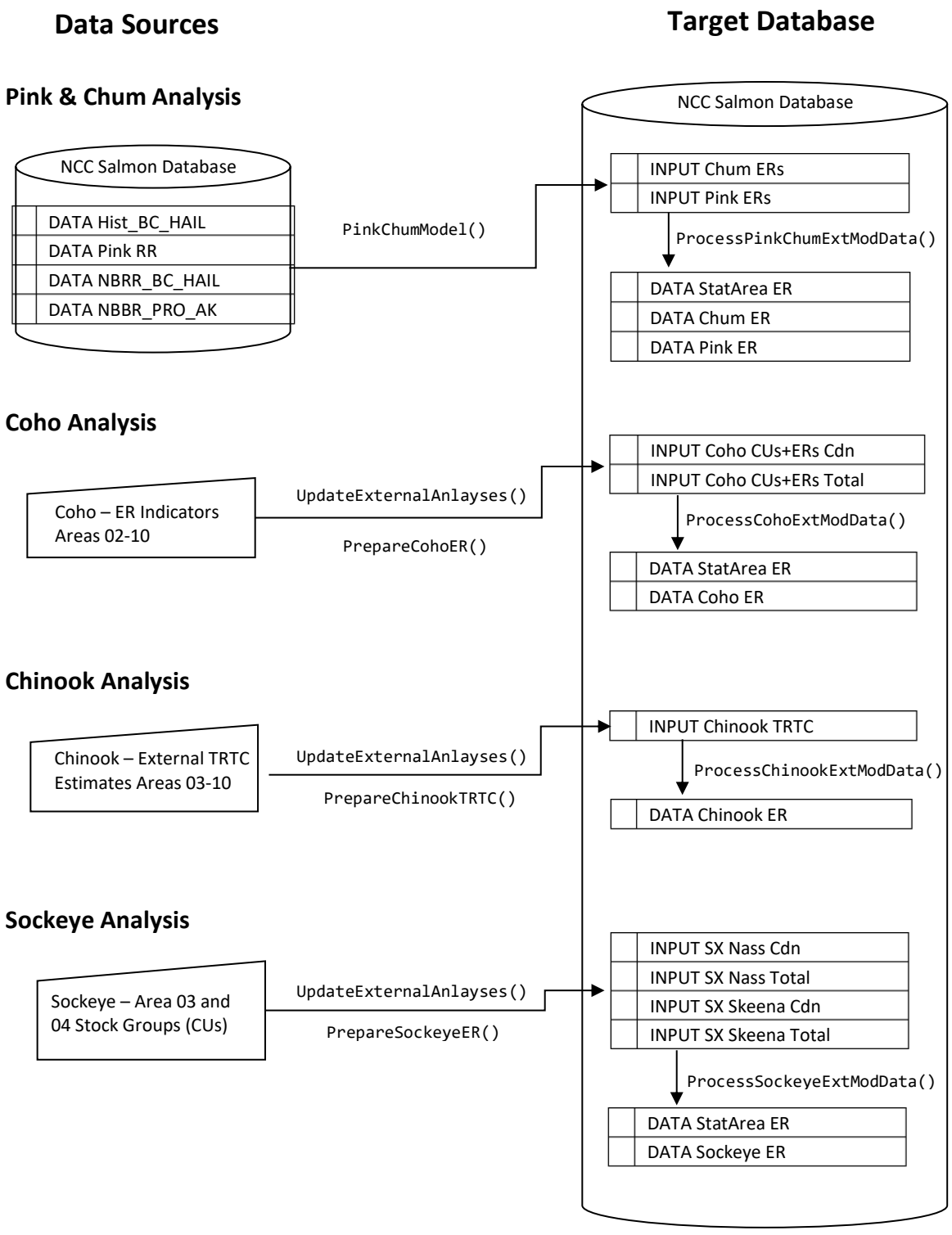

Figure 4-1. Relationship between exploitation rate estimates from species-specific models and System tables.

4.1.1 Sockeye External Analysis Exploitation Rates

The sockeye external analysis provides Canadian and total exploitation rate estimates for Nass (Table 4-2 and Table 4-3 respectively) and Skeena (Table 4-4 and Table 4-5 respectively). After which these values are used to populate the [DATA Sockeye ER] and Data exploitation rate tables (Figure 4-1) called to reconfigure and combine estimates prior to computing the final output (for further details see Section 4.3).

Table 4-2. [INPUT SX Nass Cdn ER] Canadian ER values for Nass stocks.

Field Name	Data Type	Used in Analysis	Description
Year	Integer	Yes	Calendar year
Damdochax	Number	Yes	Nass stock-specific Canadian exploitation rate
Kwinagees	Number	Yes	Nass stock-specific Canadian exploitation rate
Oweegee	Number	Yes	Nass stock-specific Canadian exploitation rate
Bowser	Number	Yes	Nass stock-specific Canadian exploitation rate
Hanna-Tin	Number	Yes	Nass stock-specific Canadian exploitation rate
MezBeach	Number	Yes	Nass stock-specific Canadian exploitation rate
BrownBear	Number	Yes	Nass stock-specific Canadian exploitation rate
Cranberry	Number	Yes	Nass stock-specific Canadian exploitation rate
Gingit+	Number	Yes	Nass stock-specific Canadian exploitation rate
Zolzap	Number	Yes	Nass stock-specific Canadian exploitation rate

Table 4-3. [INPUT SX Nass Total ER] Total ER values for Nass stocks.

Field Name	Data Type	Used in Analysis	Description
Year	Integer	Yes	Calendar year
Damdochax	Number	Yes	Nass stock-specific total exploitation rate
Kwinagees	Number	Yes	Nass stock-specific total exploitation rate
Oweegee	Number	Yes	Nass stock-specific total exploitation rate
Bowser	Number	Yes	Nass stock-specific total exploitation rate
Hanna-Tin	Number	Yes	Nass stock-specific total exploitation rate
MezBeach	Number	Yes	Nass stock-specific total exploitation rate
BrownBear	Number	Yes	Nass stock-specific total exploitation rate
Cranberry	Number	Yes	Nass stock-specific total exploitation rate
Gingit+	Number	Yes	Nass stock-specific total exploitation rate
Zolzap	Number	Yes	Nass stock-specific total exploitation rate

Table 4-4. [INPUT SX Skeena Cdn ER] Canadian ER values for Skeena stocks.

Field Name	Data Type	Used in Analysis	Description
Year	Integer	Yes	Calendar year
Kluatan+	Number	Yes	Skeena stock-specific Canadian exploitation rate
Motase	Number	Yes	Skeena stock-specific Canadian exploitation rate
Sustut+	Number	Yes	Skeena stock-specific Canadian exploitation rate
Bear+	Number	Yes	Skeena stock-specific Canadian exploitation rate
Slamgeesh	Number	Yes	Skeena stock-specific Canadian exploitation rate
Sicintine	Number	Yes	Skeena stock-specific Canadian exploitation rate
Babine-WE	Number	Yes	Skeena stock-specific Canadian exploitation rate
Babine-WM	Number	Yes	Skeena stock-specific Canadian exploitation rate
Babine-WL	Number	Yes	Skeena stock-specific Canadian exploitation rate
Babine-P	Number	Yes	Skeena stock-specific Canadian exploitation rate
Babine-F	Number	Yes	Skeena stock-specific Canadian exploitation rate
Swan+	Number	Yes	Skeena stock-specific Canadian exploitation rate
Bulkley+	Number	Yes	Skeena stock-specific Canadian exploitation rate
Morice+	Number	Yes	Skeena stock-specific Canadian exploitation rate
Kitwanga	Number	Yes	Skeena stock-specific Canadian exploitation rate
Zymoetz	Number	Yes	Skeena stock-specific Canadian exploitation rate
Kalum	Number	Yes	Skeena stock-specific Canadian exploitation rate
Lakelse	Number	Yes	Skeena stock-specific Canadian exploitation rate
Alastair	Number	Yes	Skeena stock-specific Canadian exploitation rate
Johnston	Number	Yes	Skeena stock-specific Canadian exploitation rate
Skeena Agg	Number	Yes	Skeena stock-specific Canadian exploitation rate

Table 4-5. [INPUT SX Skeena Total ER] Total ER values for Skeena stocks.

Field Name	Data Type	Used in Analysis	Description
Year	Integer	Yes	Skeena stock-specific total exploitation rate
Kluatan+	Number	Yes	Skeena stock-specific total exploitation rate
Motase	Number	Yes	Skeena stock-specific total exploitation rate
Sustut+	Number	Yes	Skeena stock-specific total exploitation rate
Bear+	Number	Yes	Skeena stock-specific total exploitation rate
Slamgeesh	Number	Yes	Skeena stock-specific total exploitation rate
Sicintine	Number	Yes	Skeena stock-specific total exploitation rate
Babine-WE	Number	Yes	Skeena stock-specific total exploitation rate
Babine-WM	Number	Yes	Skeena stock-specific total exploitation rate
Babine-WL	Number	Yes	Skeena stock-specific total exploitation rate
Babine-P	Number	Yes	Skeena stock-specific total exploitation rate
Babine-F	Number	Yes	Skeena stock-specific total exploitation rate
Swan+	Number	Yes	Skeena stock-specific total exploitation rate

Field Name	Data Type	Used in Analysis	Description
Bulkley+	Number	Yes	Skeena stock-specific total exploitation rate
Morice+	Number	Yes	Skeena stock-specific total exploitation rate
Kitwanga	Number	Yes	Skeena stock-specific total exploitation rate
Zymoetz	Number	Yes	Skeena stock-specific total exploitation rate
Kalum	Number	Yes	Skeena stock-specific total exploitation rate
Lakelse	Number	Yes	Skeena stock-specific total exploitation rate
Alastair	Number	Yes	Skeena stock-specific total exploitation rate
Johnston	Number	Yes	Skeena stock-specific total exploitation rate
Skeena Agg	Number	Yes	Skeena stock-specific total exploitation rate

4.1.2 Pink and Chum External Analysis Exploitation Rates

Exploitation rate estimates for Area 03, 04, and 05 Pink and Chum salmon are determined by the Pink&Chum model (Challenger et al 2018), which has been recently converted to R and is run as part of the main update script. Chum Area 03, 04, and 05 harvest rates are derived from the Chum Model are stored in [INPUT Chum HR] (Table 4-6) for reference purposes. The final Pink and Chum exploitation rate estimates are stored in [INPUT Pink ERs] (Table 4-7) and [INPUT Chum ERs] (Table 4-8) respectively.

Table 4-6. [INPUT Chum HR] External estimates for Chum Canadian harvest rates.

Field Name	Data Type	Used in Analysis	Description
Year	Integer	Yes	Calendar year
Area3HR	Number	Yes	Canadian harvest rate for statistical area 03
Area4HR	Number	Yes	Canadian harvest rate for statistical area 04
Area5HR	Number	Yes	Canadian harvest rate for statistical area 05

Table 4-7. [INPUT Pink ERs] External estimates Pink salmon Canadian ERs and Total ERs for Statistical Areas 03-05.

Field Name	Data Type	Used in Analysis	Description
ID	Integer	No	Unique record identifier
Year	Integer	Yes	Calendar year
A3 CDN ER	Number	Yes	Canadian exploitation rate for statistical area 03
A3 TOTAL ER	Number	Yes	Total exploitation rate for statistical area 03
A4 CDN ER	Number	Yes	Canadian exploitation rate for statistical area 04
A4 TOTAL ER	Number	Yes	Total exploitation rate for statistical area 04
A5 CDN ER	Number	Yes	Canadian exploitation rate for statistical area 05
A5 TOTAL ER	Number	Yes	Total exploitation rate for statistical area 05

Table 4-8. [INPUT Chum ERs] External estimates of Chum salmon Canadian ERs and Total ERs.

Field Name	Data Type	Used in Analysis	Description
ID	Integer	No	Unique record identifier
Year	Integer	Yes	Calendar year
A3 CDN ER	Number	Yes	Canadian exploitation rate for statistical area 03
A3 TOTAL ER	Number	Yes	Total exploitation rate for statistical area 03
A4 CDN ER	Number	Yes	Canadian exploitation rate for statistical area 04
A4 TOTAL ER	Number	Yes	Total exploitation rate for statistical area 04
A5 CDN ER	Number	Yes	Canadian exploitation rate for statistical area 05
A5 TOTAL ER	Number	Yes	Total exploitation rate for statistical area 05

4.1.3 Coho External Analysis Exploitation Rates

Similarly, for Coho salmon, two tables (Table 4-9 and Table 4-10) are used to provide the Canadian and Total ERs for each statistical area, group of statistical areas or stock group in Areas 03 and 04.

Table 4-9. [INPUT Coho CUs+ERs Cdn] Coho Canadian ER values by stock and year.

Field Name	Data Type	Used in Analysis	Description
Year	Integer	Yes	Calendar year
Area 02E	Number	Yes	Canadian exploitation rate for statistical area 02E
Area 02W	Number	Yes	Canadian exploitation rate for statistical area 02W
Area 02WE	Number	Yes	Canadian exploitation rate for statistical area 02WE
Area 03	Number	Yes	Canadian exploitation rate for statistical area 03
Area 04	Number	Yes	Canadian exploitation rate for statistical area 04
Area 06	Number	Yes	Canadian exploitation rate for statistical area 06
Area 06-08	Number	Yes	Average Canadian exploitation rate for statistical areas 06, 07, and 08
Area 08	Number	Yes	Canadian exploitation rate for statistical area 08
Area 04-09	Number	Yes	Average Canadian exploitation rate for statistical areas 04, 05, 06, 07, 08, and 09
Area 09-10	Number	Yes	Average Canadian exploitation rate for statistical areas 09 and 10

Table 4-10. [INPUT Coho CUs+ERs Total] Coho Total ER values by stock and year.

Field Name	Data Type	Used in Analysis	Description
Year	Integer	Yes	Calendar year
Area 02E	Number	Yes	Total exploitation rate for statistical area 02E
Area 02W	Number	Yes	Total exploitation rate for statistical area 02W
Area 02WE	Number	Yes	Total exploitation rate for statistical area 02WE
Area 03	Number	Yes	Total exploitation rate for statistical area 03
Area 04	Number	Yes	Total exploitation rate for statistical area 04
Area 06	Number	Yes	Total exploitation rate for statistical area 06
Area 06-08	Number	Yes	Average Total exploitation rate for statistical areas 06, 07, and 08
Area 08	Number	Yes	Total exploitation rate for statistical area 08
Area 04-09	Number	Yes	Average Total exploitation rate for statistical areas 04, 05, 06, 07, 08, and 09
Area 09-10	Number	Yes	Average Total exploitation rate for statistical areas 09 and 10

4.1.4 Chinook External Analysis Exploitation Rates

Canadian and Total ER estimates for Area 03-10 Chinook salmon and Nass, Skeena and Babine Sockeye salmon are included in the external Total Return to Canada (TRTC) tables for each of these species-stock groups that are uploaded to the NCC Salmon Database (see Sections 4.2 and 4.3).

4.2 External Estimated TRTC Results

Extensive external analysis provides complete statistical area and conservation unit TRTC output for select species. Externally derived TRTC output Area 03-04 Sockeye derived from the sockeye Northern Boundary Run Reconstruction model (Alexander et al. 2010), Area 03-10 Chinook salmon and several Chinook CUs. The results of these analyses are organized into the format of our standard TRTC output tables so these data can be readily uploaded to the NCC Salmon Database and used to produce the final TRTC outputs by CU and statistical area (Figure 4-2).

The format for the five external TRTC tables in the current database is provided in Table 4-11. Similar to the importing of externally derived exploitation rate estimates (see Section 4.1 above and Appendix B) a general function is used to do the importing with a dedicated routine to do any final cleanup or data conversions to make the results compatible with the database (Figure 4-2).

When available, the final TRTC estimates will always be based on these external TRTC estimates (see Section 6.1 below for more details). For outputs by broodyear, these estimates are combined with the age composition data prior to producing the broodyear output tables.

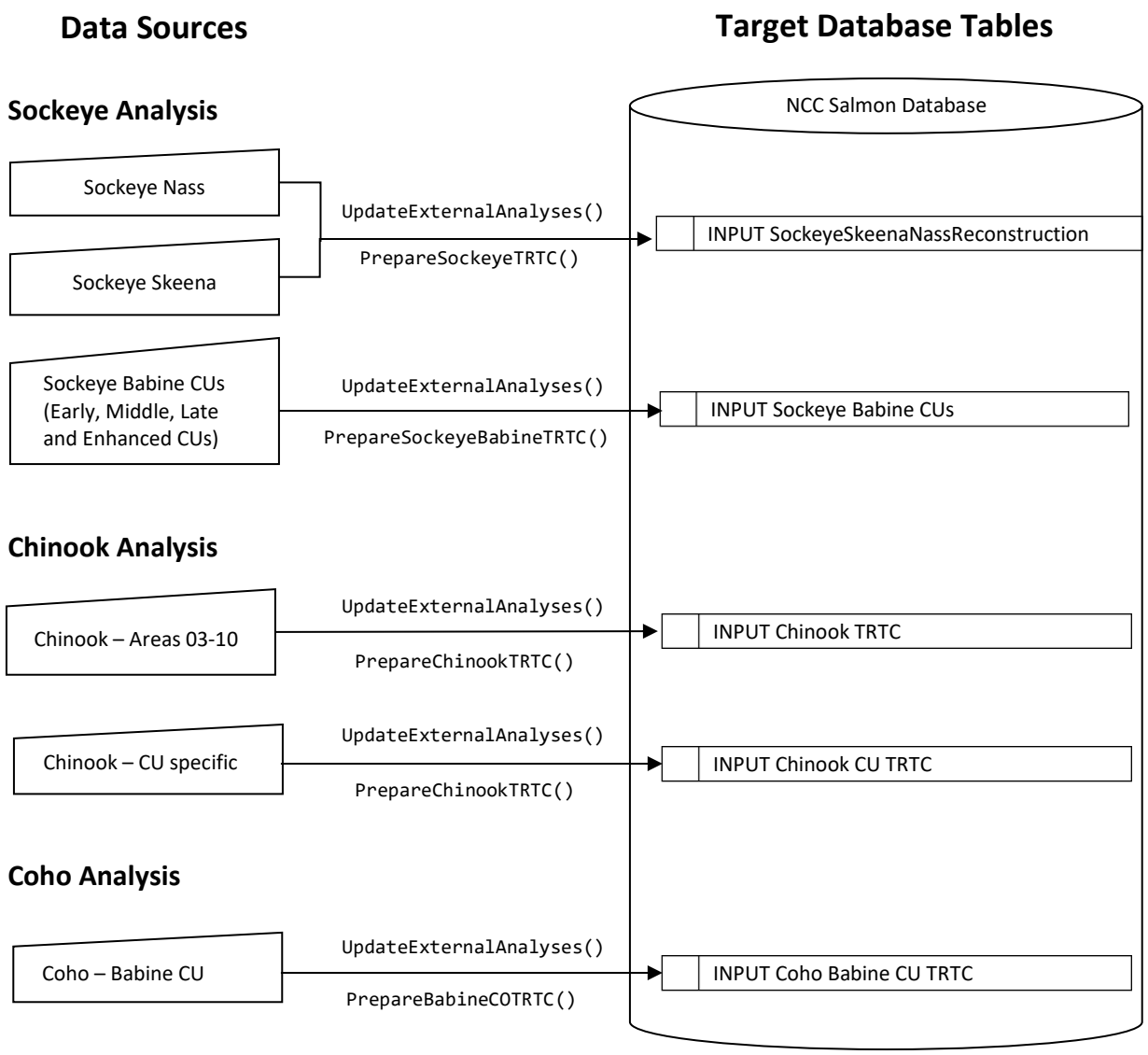

Figure 4-2. Relationships between external TRTC tables and their respective System database tables.

Table 4-11. Table structure for all external estimates of escapement, catch, ERs and TRTC for Nass, Skeena and Babine Sockeye and Chinook salmon returning to Statistical Areas 03-10 (Tables: [INPUT SockeyeSkeenaNassReconstruction], [INPUT Sockeye Babine CUs], [INPUT Chinook TRTC], [INPUT Chinook CU TRTC], and [INPUT Coho Babine CU TRTC]).

Field Name	Data Type	Used in Analysis	Description
SpeciesId	Text	Yes	Species identification code
StatArea/CU	Text	Yes	Statistical area or conservation unit code depending on the input data
CU_Name	Text	Yes	Conservation unit name – (Optional, only in table [INPUT Sockeye Babine CUs])
Year	Number	Yes	Calendar year
T_Idx_E	Number	Yes	Total index stream escapement
ExpFactor1	Number	Yes	Expansion Factor 1
AdjSum	Number	Yes	Adjusted escapement
ExpFactor2	Number	Yes	Expansion Factor 2
ObsE	Number	Yes	Observed escapement
ExpFactor3	Number	Yes	Expansion Factor 3
TE	Number	Yes	Total escapement
CDN Harvest	Number	Yes	Canadian harvest
CDN ER	Number	Yes	Canadian exploitation rate
TRTC	Number	Yes	Total Return to Canada
Total Harvest	Number	Yes	Total harvest
Total ER	Number	Yes	Total exploitation rate
Total Run	Number	Yes	Total run
Q1	Number	Yes	Indicator Streams Avg Survey Quality
Q2	Number	Yes	Indicator Streams Survey Coverage
Q3	Number	Yes	Indicator Streams % of total spawners

4.3 Post-processing of External Analysis Results

After the externally derived exploitation rate and TRTC output data has been imported into the database (i.e., see Sections 4.1 and 4.2 respectively) species-specific post-processing routines are automatically called after the data import (Figure 4-2), which then take the exploitation rate and TRTC results and populate various DATA tables used to create the TRTC output (Table 4-12). Each post processing routine completes a set tasks for each species (Table 4-13).

The [DATA StatArea ER] table (Table 4-14) contains exploitation rate estimates from non-terminal fisheries for Sockeye (i.e., Area 01, 02E, 02W, and 05), Pink and Chum (i.e., Area 03, 04, and 05), and Coho (i.e., Area 02E, 02W, 02WE, 03 – 10). Chinook statistical area ERs are contained in [INPUT Chinook TRTC] table which is directly incorporated in the TRTC output.

The species-specific DATA ER tables (e.g., [DATA Sockeye ER], [DATA Pink ER], etc.) contain statistical area specific results, averages of a range of statistical areas, and in the case of Sockeye, stock specific exploitation rate, used in conservation unit TRTC Output calculations (see Section 6.1). Estimates are mapped to specific Conservation Units by the [LOOKUP CU_ER_Indicator] table (see Section 5.2).

Table 4-12. Post-processing routines called after the import of external analysis exploitation rate and TRTC estimates.

Input Table Name	Data Type	Table Schema	Post Processing Routine	Destination Table Name	Destination Table Schema
INPUT SX Nass Cdn ER	ER	Table 4-2	ProcessSockeyeExtModData()	DATA StatArea ER	Table 4-14
INPUT SX Nass Total ER	ER	Table 4-3			
INPUT SX Skeena Cdn ER	ER	Table 4-4		DATA Sockeye ER	Table 4-15
INPUT SX Skeena Total ER	ER	Table 4-5			
INPUT Sockeye Babine CU	TRTC	Table 4-11	*No post-processing*		
INPUT SockeyeSkeenaNassReconstruction	TRTC	Table 4-11			
INPUT Pink ERs	ER	Table 4-7	ProcessPinkChumExtModData()	DATA StatArea ER	Table 4-14
INPUT Chum ERs	ER	Table 4-8		DATA Pink ER	Table 4-16
				DATA Chum ER	Table 4-17
INPUT Coho CUs+ERs Cdn	ER	Table 4-9	ProcessCohoExtModData()	DATA StatArea ER	Table 4-14
INPUT Coho CUs+ERs Total	ER	Table 4-10		DATA Coho ER	Table 4-18
INPUT Coho Babine CU TRTC	TRTC	Table 4-11	*No post-processing*		
INPUT Chinook CU TRTC	TRTC	Table 4-11	ProcessChinookExtModData()	DATA Chinook ER	Table 4-19
INPUT Chinook TRTC	TRTC	Table 4-11	*No post-processing*		

Table 4-13. Tasks completed by species-specific post-processing routines on the externally derived estimates of exploitation rate and TRTC output.

Post Processing Routine	Species	Tasks
ProcessSockeyeExtModData()	Sockeye	Transfers Nass and Skeena stock specific exploitation rate estimates from the respective INPUT tables (see Table 4-12) to [DATA Sockeye ER]. Also creates the "Hanna-Tin,MezBeac" stock estimates as an average of the "Hanna-Tin" and "MezBeach" stocks Canadian and total exploitation rates respectively. Also adds fixed estimates 20% for Area 01, 02E, and 02W to [DATA StatArea ER] and uses the yearly Lakelse stock estimate as an indicator for Area 05 exploitation rates.
ProcessPinkChumExtModData()	Pink Chum	Copies the Area 03 to 05 species-specific estimates from the respective INPUT tables (see Table 4-12) to [DATA StatArea ER], and to the respective species specific [DATA Pink ER] and [DATA Chum ER] tables. Note that additional statistical area values and averages of multiple statistical areas are added to [DATA Pink ER] and [DATA Chum ER] at a later point during the computation of conservation unit specific TRTC output (see Section 6.1 below for more details)
ProcessCohoExtModData()	Coho	Copies statistical area specific estimates and averages of area-specific estimates from the respective INPUT tables (see Table 4-12) to [DATA Coho ER]. Then populates [DATA StatArea ER] with the same INPUT values after first mapping the averages of area-specific estimates as an area specific estimate (i.e., Area 04-09 = Area 05, Area 06-08 = Area 07, Area 09-10 = Area 09 = Area 10).
ProcessChinookExtModData()	Chinook	Copies statistical area specific estimates from the external TRTC estimates (see Table 4-12) to [Data Chinook ER]. Also creates an estimate for the "CN_A04B" ER indicator which 10% of the "CN_A04" values.

Table 4-14. [DATA StatArea ER] Statistical Area exploitation rate estimates derived from external analysis results.

Field Name	Data Type	Used in Analysis	Description
SpeciesId	Text	Yes	Species identification code
StatArea/CU	Text	Yes	Statistical area
Year	Integer	Yes	Calendar year
ER	Number	Yes	Canadian exploitation rate
ER Total	Number	Yes	Total exploitation rate
Model	Text	Yes	Name of model used to produce the estimate

Table 4-15. [DATA Sockeye ER] Sockeye exploitation rate estimates used when computing CU specific exploitation rate estimates.

Field Name	Data Type	Used in Analysis	Description
Stock	Text	Yes	Statistical area, range of statistical areas, or Sockeye stock referenced as the ER Indicator in [LOOKUP CU] when deriving CU exploitation rates
Year	Integer	Yes	Calendar year
Cdn	Number	Yes	Canadian exploitation rate
Total	Number	Yes	Total exploitation rate

Note the 'Hanna-Tin,MezBeac' stock is an average of the Hanna-Tin and MezBeac stocks determined at the time of import.

Table 4-16. [DATA Pink ER] Pink exploitation rate estimates used when computing CU specific exploitation rate estimates.

Field Name	Data Type	Used in Analysis	Description
SpeciesId	Text	No	Species identification code
StatArea/CU	Text	Yes	Statistical area or a range of statistical areas referenced as the ER Indicator in [LOOKUP CU] when deriving CU exploitation rates
Year	Integer	Yes	Calendar year
CDN ER	Number	Yes	Canadian exploitation rate
Total ER	Number	Yes	Total exploitation rate

Table 4-17. [DATA Chum ER] Chum exploitation rate estimates used when computing CU specific exploitation rate estimates.

Field Name	Data Type	Used in Analysis	Description
SpeciesId	Text	No	Species identification code
StatArea/CU	Text	Yes	Statistical area or a range of statistical areas referenced as the ER Indicator in [LOOKUP CU] when deriving CU exploitation rates
Year	Number	Yes	Calendar year
CDN ER	Number	Yes	Canadian exploitation rate
Total ER	Number	Yes	Total exploitation rate

Table 4-18. [DATA Coho ER] Coho exploitation rate estimates used when computing CU specific exploitation rate estimates.

Field Name	Data Type	Used in Analysis	Description
Keystream	Text	Yes	Statistical area or a range of statistical areas referenced as the ER Indicator in [LOOKUP CU] when deriving CU exploitation rates
Year	Integer	Yes	Calendar year
Cdn	Number	Yes	Canadian exploitation rate
Total	Number	Yes	Total exploitation rate

Table 4-19. [DATA Chinook ER] Chinook exploitation rate estimates used when computing CU specific exploitation rate estimates.

Field Name	Data Type	Used in Analysis	Description
ER Indicator	Text	Yes	ER indicator code referenced in [LOOKUP CU] when deriving CU exploitation rates
Year	Integer	Yes	Calendar year
Cdn ER	Number	Yes	Canadian exploitation rate
Total ER	Number	Yes	Total exploitation rate

5. ANALYSIS PARAMETERS AND CORRESPONDING LOOK-UP TABLES

A variety of analysis parameters are maintained in an assortment of LOOKUP tables (Figure 5-1 and Table 5-1), which is used to generate the final output data (see Section 6). This includes indicator streams (Section 5.1), conservation unit ER indicators (Section 5.2), Expansion Factor 2 (Section 5.3 and 5.4), and Expansion Factor 3 (Section 5.5).

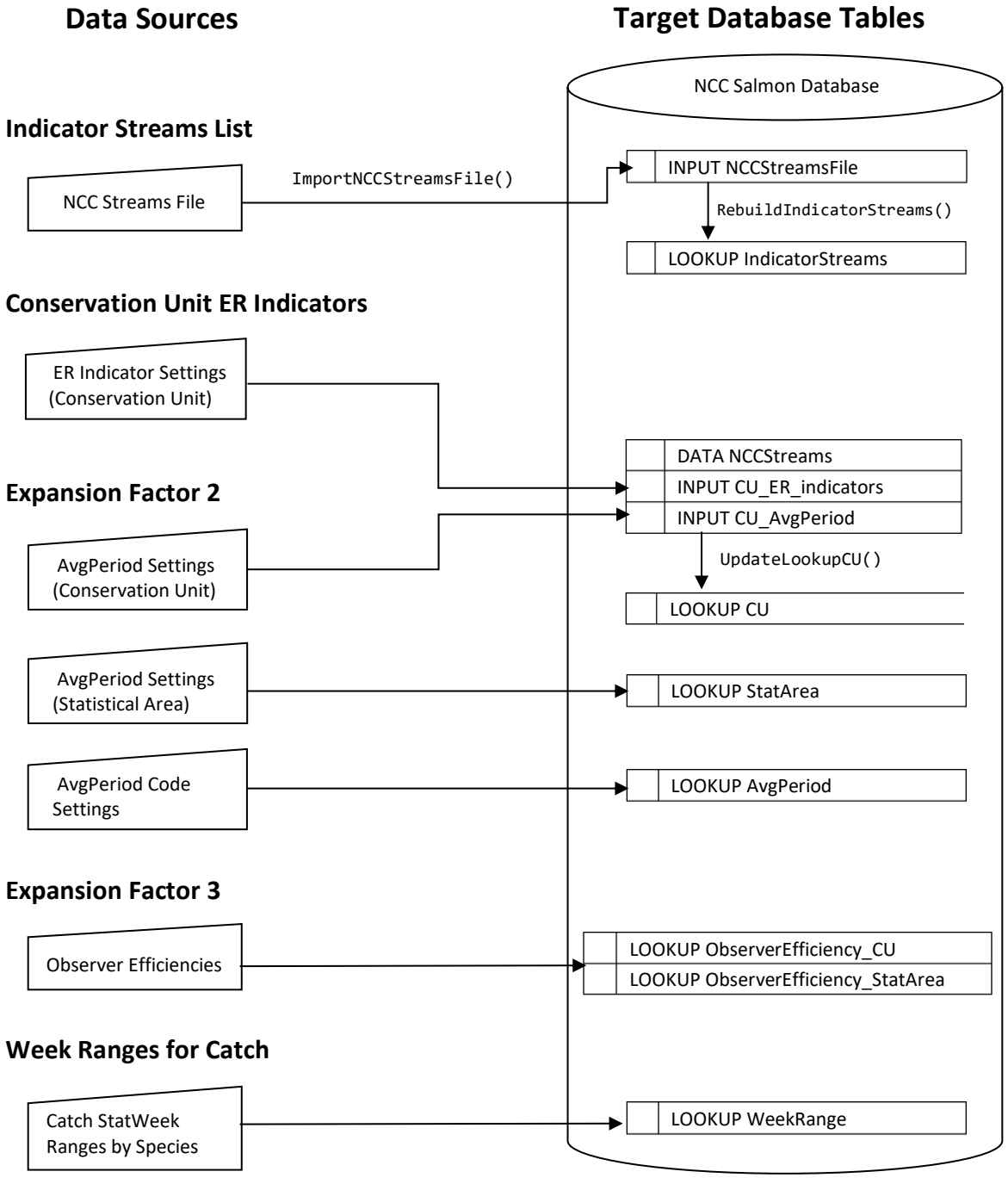

Figure 5-1. Relationship between analysis parameters and settings and the respective System tables.

Table 5-1. Summary of LOOKUP tables and associated data sources and reporting sections.

Report Section	Table Name	Table Schema	Details
5.1 NCC Streams File and Indicator Streams	INPUT NCCStreamsFile	Table 5-2	List of NCC Streams along with audit results and a field indicating whether or not the stream should be used as an indicator, also serves as the basis of the NCC Stream Escapement Output (see Section 6.3)
	LOOKUP IndicatorStreams	Table 5-3	List of indicator streams used to determine indicator stream escapement, which is based on [INPUT NCCStreamsFile]
5.2 Conservation Unit ER Indicators	INPUT CU_ER_Indicators	Table 5-4	List of exploitation rate indicators by species-specific conservation unit code. Used to populate [LOOKUP CU] when the table is rebuilt during data updates.
5.3 AvgPeriod Code for Expansion Factor 2	LOOKUP AvgPeriod	Table 5-6	Translates AvgPeriod code into a set of years used when computing Expansion Factor 2.
	INPUT CU_AvgPeriod	Table 5-7	List of AvgPeriod codes for each species-specific conservation unit code. Used to populate [LOOKUP CU] when the table is rebuilt during data updates.
	LOOKUP StatArea	Table 5-8	Statistical area specific AvgPeriod settings.
5.4 Conservation Unit and StatArea Look-up Tables	LOOKUP CU	Table 5-9	Conservation unit settings and summaries.
	LOOKUP StatArea	Table 5-8	Statistical area specific settings.
5.5 Observer Efficiencies	LOOKUP ObserverEfficiency_CU	Table 5-10	Observer efficiencies used for expansion factor 3 when computing conservation unit specific escapement.
	LOOKUP ObserverEfficiency_StatArea	Table 5-11	Observer efficiencies used for expansion factor 3 when computing statistical area specific escapement.
5.6 Reference Period Look-up Tables	LOOKUP ExpFactor1_CU	Table 5-13	Expansion Factor 1 reference periods for conservation units
	LOOKUP ExpFactor1_StatArea	Table 5-14	Expansion Factor 1 reference periods for statistical areas
	LOOKUP ExpFactor2_CU	Table 5-15	Expansion Factor 2 conservation unit settings and values
	LOOKUP ExpFactor2_StatArea	Table 5-16	Expansion Factor 2 statistical area settings and values
5.7 Miscellaneous Settings	LOOKUP Species_List	Table 5-17	List of species names and associated species codes
	LOOKUP WeekRange	Table 5-18	StatWeeks used when computing species specific catch. Currently, no StatWeek limitations have been specified.

5.1 NCC Streams File and Indicator Streams

The full list of NCC streams along with meta data such as in the indicator status has been maintained in a "NCC Streams File." Periodically, this file has been used by DFO to update an "indicator stream" field in the NuSEDS database. The procedure for approving and conducting regular updates to the indicator stream field in the NuSEDS database is not currently established, so until then, the list of indicator streams used in the System will be the latest version of the "NCC Streams File" file as approved by the senior North Coast stock assessment biologist.

Externally, the "NCC Streams File" usually contains yearly escapement estimates for each stream, however these values are removed before storing all other data in the [INPUT NCC StreamsFile] data table (Table 5-2). The most up-to-date escapement estimates are then added back on as part of the output data (see Section 6.3). The source column in Table 5-2 indicates the origin of the information in each data field. LGL refers to work conducted by LGL Limited from 2003-2013. Holtby refers to information provided by Blair Holtby in a dataset he provided to LGL dated 25 October 2011.

The data stored in the [INPUT NCC StreamsFile] is also used to populate the [LOOKUP IndicatorStreams] table which contains a full list of the indicator streams, and relevant meta data (e.g., survey quality), is used to generate output data (Table 5-3). This update is handled by the `RebuildIndicatorStreams()` routine in the NCCSDB package (Appendix C), which is called by the main analysis script (Appendix B).

Table 5-2. [INPUT NCCStreamsFile] Data fields associated with the import of the NCC Streams file.

Field Name	Data Type	Used in Analysis	Description	Source
ID	Integer	No	Unique record identifier	LGL
POP_ID	Integer	Yes	Blair POP_ID(100,000 added to Pink odd year PopIds, 200,000 added to Pink even year PopIds)	Holtby
SPP	Text	Yes	Species identification code, this uses the "SPECIES_QUALIFIED" format found in the Conservation Unit Site List database that accompanies NuSEDS data releases. This code differs from SpeciesId code used in the NCC Salmon Databse in that Sockeye are broken down into river (SER) and lake (SEL) types. Use the ConvertSpeciesQualified() function to get to the standard SpeciesId format used across the NCC Salmon Database.	Holtby
GFE_ID	Integer	No	NuSEDS population identifier	Holtby
SYS_NM	Text	Yes	Stream name	Holtby
source	Text	No	Data source (usually NuSEDS)	Holtby
popMAP	Integer	No	POP_ID to be used for these data	Holtby
EXPN	Number	No	Proposed expansion factor	Holtby

Field Name	Data Type	Used in Analysis	Description	Source
yLAT	Number	No	Stream location – latitude	Holtby
xLONG	Number	No	Stream location – longitude	Holtby
faz_acro	Text	No	Freshwater Adaptive Zone	Holtby
maz_acro	Text	No	Marine Adaptive Zone	Holtby
jaz_acro	Text	No	Joint Adaptive Zone	Holtby
CU_fname	Text	No	Species+CU name	Holtby
CU_facro	Text	No	Species+CU acro	Holtby
CU_findex	Text	No	Species+CU index	Holtby
CU_name	Text	No	CU name	Holtby
CU_acro	Text	No	CU acronym	Holtby
CU_index	Text	Yes	A code assigned to the CU that when prefixed by the species code becomes the CU_index. Originally, these codes were assigned in a roughly S to N geographic order. With each revision to the CUs that order has been lost.	Holtby
Area	Text	Yes	Statistical Area using the NuSEDS format (i.e., 1 to 10 without the leading zero). Use Convert2StatArea() to get the NCC Salmon Database format.	Holtby
SEP_ENH	Text	No	Enhancement flag (Y=enhanced)	Holtby
IsIndicator	Text	Yes	Indicator stream ('Y'=Yes)	Holtby
IsInFiltEsc	Text	No	Escapement filter flag (blank in this dataset)	Holtby
nrecs	Text	No	Old- Number of NuSEDS records for this POPID	Holtby
nins	Text	No	Old- Number of stream inspections	Holtby
npres	Text	No	Old- Number of records indicating fish presence	Holtby
nnumest	Text	No	Old- Number of numerical estimates	Holtby
pins_rec	Number	No	Old- # inspections records/# NuSEDS records	Holtby
ppres_ins	Number	No	Old- # presence records/# inspection records	Holtby
pest_pres	Number	No	Old- # numerical estimates/# presence records	Holtby
fr_timing	Text	No	Fraser run timing (NA for this dataset)	Holtby
OL_GRP_NM	Text	No	Group of statistical areas	Holtby
Fraser_mnemonic	Text	No	Fraser timing group (NA for this dataset)	Holtby
CD_findex-name	Text	No	Index name (Blank for this dataset)	Holtby
StatArea	Text	No	Statistical area code	LGL
Reviewer	Text	No	Last name of reviewer for indicator streams	LGL
QA	Integer	Yes	Survey quality assessment code	LGL
Method	Text	No	Survey method (incomplete)	LGL
Race	Text	No	Run timing (incomplete)	LGL
WildCode	Text	No	% Wild code (incomplete)	LGL
WildRigor	Text	No	% Wild – Data quality code (incomplete)	LGL
CSAR-Meth	Text	No	Core Stock Assessment Review method	LGL
Last8yrs	Text	No	Number of years with data between 2007-2014	LGL
CSAR2006	Text	No	2006 Core Stock Assessment Review indicator	LGL

Field Name	Data Type	Used in Analysis	Description	Source
			streams ('Y'=Yes)	
50s	Integer	No	Number of non-zero escapement years in the 1950s	LGL
60s	Integer	No	Number of non-zero escapement years in the 1960s	LGL
70s	Integer	No	Number of non-zero escapement years in the 1970s	LGL
C80s	Number	No	Number of non-zero escapement years in the 1980s	LGL
C90s	Number	No	Number of non-zero escapement years in the 1990s	LGL
C00s	Number	No	Number of non-zero escapement years in the 2000s	LGL
Avg1950-70	Number	No	Average of non-zero escapement estimates for 1950-1070	LGL
Avg1980-1999	Number	No	Average of non-zero escapement estimates for 1980-2010	LGL
Avg2000-14	Number	No	Average of non-zero escapement estimates for 1980-2010	LGL
Review2015	Text	No	Survey method from 2015 review	LGL
Reviewer1	Text	No	Name of 2015 reviewer	LGL
Group	Text	No	Group responsible for conducting the escapement surveys	LGL
Comments	Text	No	Comments	LGL
Comments1	Text	No	Comments	LGL

Table 5-3. [LOOKUP IndicatorStreams] List of escapement indicator streams and the NuSEDS identifier (PopId), the CU and assessment code for the average quality of the escapement data for each indicator stream (QA).

Field Name	Data Type	Used in Analysis	Description
SpeciesId	Text	No	Species identification code
StreamName	Text	No	Stream name
IndexId	Text	Yes	Unique stream by species identifier (combination of SpeciesID and PopId fields). Used internally instead of PopId in order to deal with Pink even and odd year populations that have the same PopId
StatArea	Text	No	Statistical area
CU	Text	No	Conservation unit code
QA	Text	Yes	Survey quality assessment code
PopId	Integer	No	NuSEDS POP_ID field. Unique number used to identify a specific population of salmon defined by species, spawning stream and run-timing

5.2 Conservation Unit ER Indicators

When computing the conservation unit TRTC output (see Section 6.1) the best available exploitation rate estimate is used for each CU. The species-specific exploitation rate estimates used are kept in the species-specific DATA ER table which contain exploitation rate estimates for a specific statistical area, group of statistical areas or stock group (see Table 4-12 and Table 4-13 is Section 4.3). Each CU is then linked the respective DATA ER record via CU ER Indicator, which is the best available set of annual ER estimates for that CU contained in the [INPUT CU_ER_Indicators] (Table 5-4). For convenience purposes the [INPUT CU_ER_Indicators] records are not directly referenced, but rather are copied over to the [LOOKUP CU] table when it is updated (see Section 5.4). This allows ER indicators to be modified in the [LOOKUP CU] table so that the impact on calculations can be determined, yet allows for the values to be reverted to the original settings retained in the [INPUT CU_ER_Indicators] table. The sources from the full list of ER indicators comes from a series of files (Table 5-5).

Table 5-4. [INPUT CU_ER_Indicators] CUs and the corresponding ER indicator statistical area, statistical area range, or stock groups.

Field Name	Data Type	Used in Analysis	Description
SpeciesId	Text	Yes	Species identification code
CU	Text	Yes	Conservation unit code
CU_Name	Text	No	Conservation unit name
ER_Indicator	Text	Yes	Conservation unit exploitation rate indicator code which designates the statistical area, range of statistical areas or stock that should be used as an exploitation rate proxy

Table 5-5. Source files containing exploitation rate indicators by CU for each species of salmon.

Species	Source File	Source File Year
Sockeye	*NCCC Salmon Tables2-6 CUs+ERs_28Nov2011_kke.xls*	2011
Pink Chum	*Area 3-4-5 Pink & Chum HR Model 1954-2010_21Jan2013.xlsx*	2013
Coho	*Coho_CUs+ERs_1980-10_22Nov2011_forTony.xlsx*	2011
Chinook	*Chinook_CUs+HRs_1980-07SC_20091027.xls*	2009

5.3 AvgPeriod Code for Expansion Factor 2

The AvgPeriod code is used when computing decadal averages for Expansion Factor 2. An AvgPeriod code of 0 indicates that decadal averages should be used as computed without modification. For AvgPeriod code values greater than zero indicates that a reference period should be used in place for each decadal average, with the settings stored in the [LOOKUP AvgPeriod] table (Table 5-6). Currently, the codes for the averaging period are: 1 = 1980s, 2 = 1990s, 3=2000s, 4=1980-1999 and 5=all analysis years. By storing the year to AvgPeriod code

associations in the [LOOKUP AvgPeriod] table, non-contiguous years can be easily programmed if desired.

The AvgPeriod code settings for each CU is stored in the [INPUT CU_AvgPeriod] table (Table 5-7). For convenience purposes the [INPUT CU_AvgPeriod] records are not directly referenced, but rather are copied over to the [LOOKUP CU] table when it is updated (see Section 5.4). This allows AvgPeriod codes to be modified in the [LOOKUP CU] table (e.g., to test out the impact on output results), yet allows for the values to be reverted to be quickly reverted to original settings.

The AvgPeriod code settings for each statistical area by species is stored in the [LOOKUP StatArea] table (Table 5-8).

Table 5-6. [LOOKUP AvgPeriod] Table defining the decade each year belongs to, and the periods used to calculate the averages of the escapement estimates for each year.

Field Name	Data Type	Used in Analysis	Description
Year	Integer	Yes	Calendar year
AvgPeriod1	Logical	Yes	Indicates whether year is included for AvgPeriod code = 1
AvgPeriod2	Logical	Yes	Indicates whether year is included for AvgPeriod code = 2
AvgPeriod3	Logical	Yes	Indicates whether year is included for AvgPeriod code = 3
AvgPeriod4	Logical	Yes	Indicates whether year is included for AvgPeriod code = 4
AvgPeriod5	Logical	Yes	Indicates whether year is included for AvgPeriod code = 5

Table 5-7. [INPUT CU_AvgPeriod] AvgPeriod code for each species-specific conservation unit

Field Name	Data Type	Used in Analysis	Description
SpeciesId	Text	Yes	Species identification code
CU_Name	Text	No	Conservation unit name
CU_Acro	Text	No	Conservation unit acronym
CU_Index	Text	No	Conservation unit index code
CU	Text	Yes	Conservation unit code
StatArea	Text	No	Statistical area
AvgPeriod	Integer	Yes	AvgPeriod code for expansion factor 2

Table 5-8. [LOOKUP StatArea] Table of species-specific statistical area settings, which currently only includes AvgPeriod.

Field Name	Data Type	Used in Analysis	Description
SpeciesId	Text	Yes	Species identification code
StatArea	Text	Yes	Statistical area
AvgPeriod	Integer	Yes	Average period code used when computing expansion factor 2

5.4 Conservation Unit and StatArea Look-up Tables

Information and analysis parameters associated with statistical areas and conservation units are kept in the [LOOKUP StatArea] and [LOOKUP CU] tables (Table 5-8 and Table 5-9 respectively). Both tables are species-specific as area settings can depend on the salmon species under consideration and conservation units defined within the System are also species dependent.

While [LOOKUP StatArea] currently only contains AvgPeriod settings (see Section 5.3), the [LOOKUP CU] table contains additional information such as the CU name, ER indicator (see Section 5.2) and summaries of the total number of available streams, indicator streams and survey rating qualities (Table 5-9). Ratings for survey quality were determined on a five point scale:

1. **Poor quality** – An estimate of poor reliability due to few surveys, counting deficiencies, etc.;
2. **Fair quality** – An estimate of moderate reliability based on two or more visual inspections (i.e., low quality AUC estimate);
3. **Good quality** – An estimate of good reliability based on three or more visual inspections (i.e., medium quality AUC estimate);
4. **Very Good quality** – An estimate of high reliability based on MR data, almost complete fence counts, or high quality AUC estimates;
5. **Excellent quality** – An estimate of very high reliability from an unbreached fence count.

The [LOOKUP CU] table is also rebuilt based on content [DATA NCCStreams], [INPUT CU_ER_Indicators], and [INPUT CU_AvgPeriod] tables (Figure 5-1). As such this table tends to be rebuilt only after all the other data tables have been appropriately updated (see Appendix B for further details).

Table 5-9. [LOOKUP CU] Conservation unit information about CUs and the periods (years) used to compute average escapements.

Field Name	Data Type	Used in Analysis	Description
SpeciesId	Text	Yes	Species identification code
CU	Text	Yes	Conservation unit code
CU_Name	Text	Yes	Conservation unit name
CU_Name_old	Text	No	Previous conservation unit name
CU_Acro	Text	No	Conservation unit acronym
CU_Acro_old	Text	No	Previous conservation unit acronym
AvgPeriod	Integer	Yes	Average period code used when computing expansion factor 2
AvgPeriodYears	Text	No	Summary of years specified by the average period code
ER_Indicator	Text	Yes	Conservation unit exploitation rate indicator code which designates the statistical area, range of statistical areas or stock that should be used as an exploitation rate proxy
StatArea	Text	No	Statistical area, range of statistical areas associated with a CU

Field Name	Data Type	Used in Analysis	Description
StatArea_old	Text	No	Previous statistical area, range of statistical areas associated with a CU
TotalStreams	Integer	Yes	Total number of streams in the conservation unit.
IndicatorStreams	Integer	Yes	Total number of indicator streams found in the conservation unit
QA1	Integer	Yes	Count of indicators with a survey quality assessment rating of 1
QA2	Integer	Yes	Count of indicators with a survey quality assessment rating of 2
QA3	Integer	Yes	Count of indicators with a survey quality assessment rating of 3
QA4	Integer	Yes	Count of indicators with a survey quality assessment rating of 4
QA5	Integer	Yes	Count of indicators with a survey quality assessment rating of 5

5.5 Observer Efficiencies (i.e., Expansion Factor 3)

Expansion Factor 3 values for each species by conservation units and statistical areas are stored in the [LOOKUP ObserverEfficiency_CU] (Table 5-10) and [LOOKUP ObserverEfficiency_StatArea] (Table 5-11) tables respectively.

Table 5-10. [LOOKUP ObserverEfficiency_CU] Observer efficiency estimates (i.e., Expansion Factor 3) for each combination of species, CU and decade.

Field Name	Data Type	Used in Analysis	Description
SpeciesId	Text	Yes	Species identification code
StatArea/CU	Text	Yes	Conservation unit code
Decade	Text	Yes	The decade to which expansion factor 3 value applies
ExpFactor3	Number	Yes	The expansion factor 3 value

Table 5-11. [LOOKUP ObserverEfficiency_StatArea] Observer efficiency estimates (i.e., Expansion Factor 3) for each combination of species, statistical area and decade.

Field Name	Data Type	Used in Analysis	Description
SpeciesId	Text	Yes	Species identification code
StatArea/CU	Text	Yes	Statistical area
Decade	Text	Yes	The decade to which expansion factor 3 value applies
ExpFactor3	Number	Yes	The expansion factor 3 value

5.6 Reference Period Look-up Tables

Estimates of escapement to statistical areas and conservation units rely on the determination of three expansion factors, two of which (i.e., Expansion Factor 1 and Expansion Factor 2) rely on computing decadal averages. When decadal averages cannot be computed due to lack of escapement data a reference value, which may be another decade or a larger span of time, is

used instead. These calculations are done internally within the R routines contained within the custom NCCSDB R-package (see Appendix B and Appendix C), but for the purpose of transparency the results have been saved to a series of LOOKUP tables (Table 5-12).

Reference values for Expansion Factor 1 are determined by an algorithm within the `ExpFactor1RefDecade()` routine that first looks for the closest decade with complete information, then falls back to a series of reference periods that includes a larger span of time (see Table 5-12 for further details). The Expansion Factor 1 decadal references, as determined by the `ExpFactor1RefDecade()` routine, are stored in the [LOOKUP ExpFactor1_CU] (Table 5-13) and [LOOKUP ExpFactor1_StatArea] (Table 5-14) tables respectively for conservation units and statistical areas calculations.

Reference values used for Expansion Factor 2 are based on the AvgPeriod codes stored in the [LOOKUP StatArea] (Table 5-8) and [LOOKUP CU] (Table 5-9) tables (see Section 5.3). The results from all Expansion Factor 2 calculations for conservation units and statistical areas are stored in [LOOKUP ExpFactor2_CU] (Table 5-15) and [LOOKUP ExpFactor2_StatArea] (Table 5-16) for conservation units and statistical areas respectively. This includes conservation units or statistical areas using the default approach (i.e., an AvgPeriod = 0) as well as those using a non-default values.

Table 5-12. List of reference period tables.

Table Name	Table Schema	Details
LOOKUP ExpFactor1_CU	Table 5-13	When calculating Expansion Factor 1, there must be data for all index streams within a statistical area or conservation unit for each decade. In the case of insufficient data for a decade, a reference period is used in its place. Where available the closest decade with observations for all indicator streams is used. If no suitable reference decade is available alternate reference periods are used in the following order: 1) 1980s and 1990s (i.e., 1980-1999); 2) 1980s, 1990s and 2000s (i.e., 1980-2009); and 3) all years. If any indicator stream has no observations with the reference period the next substitute period is checked. The final default (all years) computes averages across all specified analysis years.
LOOKUP ExpFactor1_StatArea	Table 5-14	All species-specific conservation units or statistical areas decades that required a reference value are listed in theses tables along with the reference period that was used. If no suitable replacement decade is found a warning will be issued indicating the species and conservation unit or statistical area results that are affected. In this case, the reference period averages (i.e., all years) will be used as is when computing Expansion Factor 1.
LOOKUP ExpFactor2_CU LOOKUP ExpFactor2_StatArea	Table 5-15 Table 5-16	Decadal stream Expansion Factor 2 values computed for all conservation units or statistical areas. The AvgPeriod code used when computing Expansion Factor 2 is also indicated.

Table 5-13. [LOOKUP ExpFactor1_CU] Expansion factor1 reference "decades" for conservation units where there were insufficient streams available within a decade.

Field Name	Data Type	Used in Analysis	Description
SpeciesId	Text	Yes	Species identification code
CU	Text	Yes	Conservation unit code
Decade	Text	Yes	Expansion factor 1 decade
DecadeNo	Number	No	Expansion factor 1 decade code
TotalStreams	Integer	Yes	Total number of indicator streams in the CU
Available	Integer	Yes	Number of available indicator streams in the CU
ReplacementDecade	Text	Yes	Expansion factor 1 decade to use as a replacement
ReplacementDecadeNo	Integer	No	Expansion factor 1 decade code to use as a replacement

Table 5-14. [LOOKUP ExpFactor1_StatArea] Expansion factor1 reference "decades" for statistical areas where there were insufficient streams available within a decade.

Field Name	Data Type	Used in Analysis	Description
SpeciesId	Text	Yes	Species identification code
StatArea	Text	Yes	Statistical area
Decade	Text	Yes	Expansion factor 1 decade
DecadeNo	Number	No	Expansion factor 1 decade code
TotalStreams	Integer	Yes	Total number of indicator streams in the statistical area
Available	Integer	Yes	Number of available indicator streams in the statistical area
ReplacementDecade	Text	Yes	Expansion factor 1 decade to use as a replacement
ReplacementDecadeNo	Integer	No	Expansion factor 1 decade code to use as a replacement

Table 5-15. [LOOKUP ExpFactor2_CU] Expansion factor 2 decadal values for conservation units where the AvgPeriod code indicated the use of a reference period.

Field Name	Data Type	Used in Analysis	Description
SpeciesId	Text	Yes	Species identification code
CU	Text	Yes	Conservation unit code
Decade	Text	Yes	Expansion factor 2 decade
AvgPeriod	Number	Yes	Expansion factor 2 AvgPeriod code used, missing value indicates that the current decade should be used.
ExpFactor2	Number	Yes	Expansion factor 2 value

Table 5-16. [LOOKUP ExpFactor2_StatArea] Expansion factor 2 decadal values for statistical areas where the AvgPeriod code indicated the use of a reference period.

Field Name	Data Type	Used in Analysis	Description

SpeciesId	Text	Yes	Species identification code
StatArea	Text	Yes	Statistical area
Decade	Text	Yes	Expansion factor 2 decade
AvgPeriod	Number	Yes	Expansion factor 2 AvgPeriod code used, missing value indicates that the current decade should be used.
ExpFactor2	Number	Yes	Expansion factor 2 value

5.7 Miscellaneous Settings

Mapping between the species codes found in the NuSEDS database and the System species code (see Table 3-3) is stored in the [LOOKUP Species_List] table (Table 5-17). Calculations of total catch can also be limited by StatWeeks, by species area and gear type based on settings stored in the [LOOKUP WeekRange] table (Table 5-18). While these settings were important to previous analyses, the current settings for minimum and maximum weeks do not exclude any catch data from analyses.

Table 5-17. [LOOKUP Species_List] Table mapping the abbreviated name of the species (used as an identifier throughout the System) to a full name for the menu.

Field Name	Data Type	Used in Analysis	Description
SpeciesId	Text	Yes	Species identification code
SpeciesName	Text	Yes	Species identification name found in NuSEDS

Table 5-18. [LOOKUP WeekRange] Defines the statistical week range (minimum to maximum) for which catch data is valid, by species, statistical area and gear.

Field Name	Data Type	Used in Analysis	Description
SpeciesId	Text	Yes	Species identification code
StatArea	Text	Yes	Statistical area
GearId	Text	Yes	Gear type (i.e., GN = gillnet, SN = seine net, TR = trawl, UN = unknown)
Min_StatWeek	Integer	Yes	Minimum statistical week of catch to include in total catch calculations
Max_StatWeek	Integer	Yes	Maximum statistical week of catch to include in total catch calculations

Note: Primary importance has been in previous analyses. Current settings for minimum and maximum weeks do not exclude any catch data from the analysis.

6. GENERATING OUTPUT DATA

All output results are generated as part of the main analysis script (Appendix B) and are saved both as external Microsoft Excel files and as OUTPUT tables within the System (Figure 2-1). The relationships between tables used as input data and corresponding output tables and NCCSDB routines are summarized in Figure 6-1 and described further in Appendix B.

The TRTC results (Section 6.1) and Age Tables (Section 6.2) are typically the primary output of interest. The updated “NCC Streams File” output is covered in Section 6.3

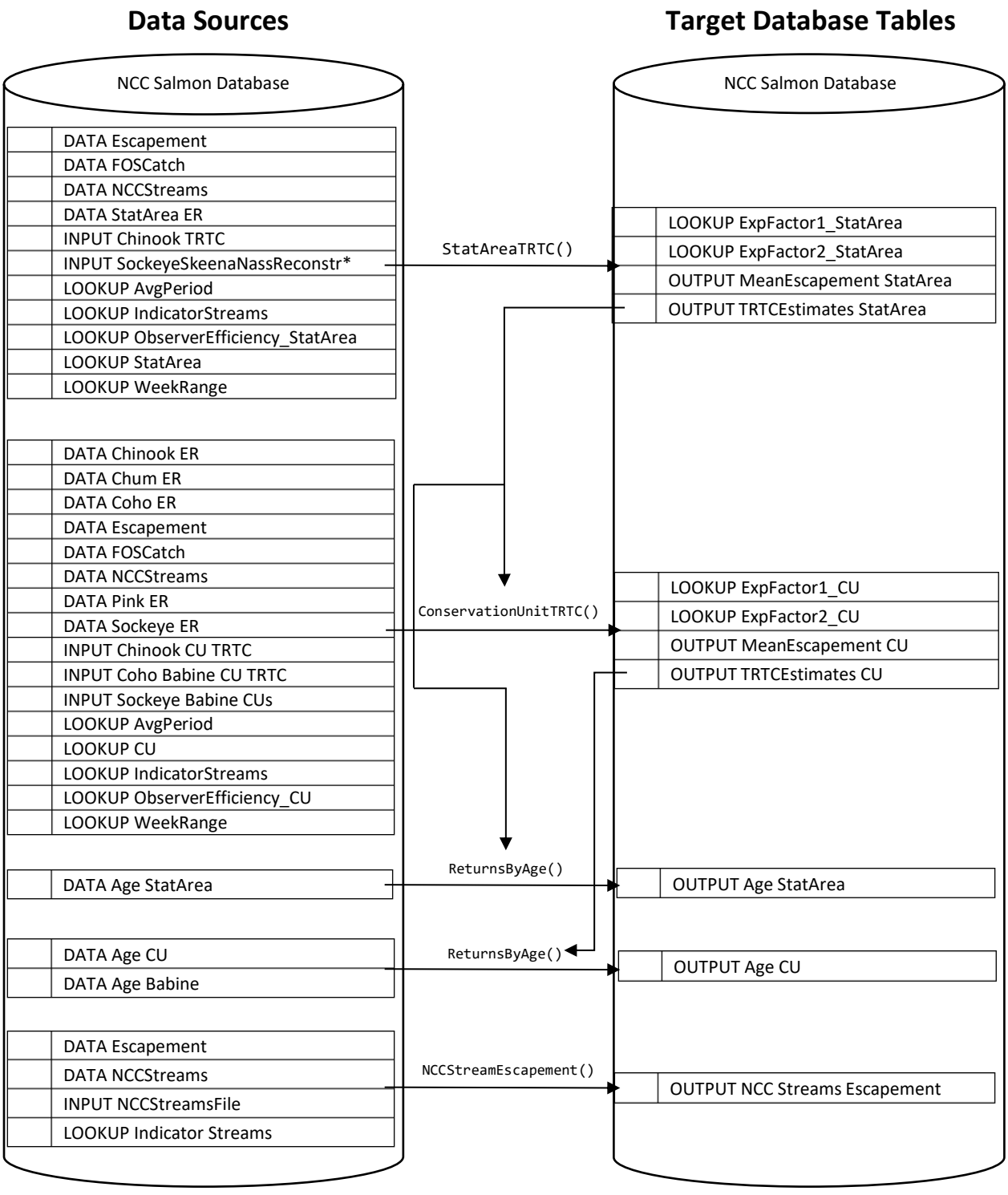

Figure 6-1. Flow diagram of the data tables and associated routines used to produce output results.

6.1 TRTC Output

TRTC tables for statistical areas are generated as part of the main analysis and generate estimates based on the deriving estimates of Canadian exploitation rates from Total Canadian Catch (TCC) and escapement (E) estimates where ER= TCC/(TCC+E) (herein referred to as the TCC&E model). Estimates from the TCC&E model are used for generating statistical area estimates for areas with a terminal fishery. For non-terminal fisheries ER estimates or complete ER estimates are derived from external model estimates (see Section 4.1 and 4.2 respectively). Estimates of harvest, catch and the total run are then back calculated based on the externally derived ER estimates and total escapement estimates. Statistical area TRTC output for all species and statistical areas are then saved to the [OUTPUT TRTCEstimates StatArea] table (Table 6-1).

Conservation unit TRTC output is based on statistical area ER estimates and external model ER estimates as well as the ER indicators defined for each conservation unit (see Section 5.2). Estimates of harvest, catch and the total run are then calculated based on the indicator ER value and the total escapement estimate. TRTC output for all species and all conservation units with indicator streams (see Section 5.1) are then saved to the [OUTPUT TRTCEstimates CU] table (Table 6-1).

In both cases, total escapement to a given geographical area (i.e., statistical area or conservation unit) are computed based on indicator streams identified within the geographical area that best represent the salmon population. The total indicator escapement (T_Idx_1) is then expanded via the Expansion Factor 1 value to account for any indicator streams escapement that may have been missing in a given year. This adjusted escapement estimate (AdjSum) is then expanded, by Expansion Factor 2, to account for the contribution of streams within the geographical area that are not indicator streams to produce the observed escapement (Obs). Observed escapement is then expanded one last time, by Expansion Factor 3, to account for the fact that observed escapement will underrepresent the total escapement to the area (i.e., one never observes 100% of the total escapement).

Finally, as part of these calculations (i.e., Expansion Factors 2) the mean decadal escapement by stream is computed and may be of general interest to investigators. As a result, the mean decadal escapement by stream is also saved to the [OUTPUT MeanEscapement] table for transparency and reference purposes (Table 6-2).

Table 6-1. Shared output format for conservation unit and statistical area escapement, catch and TRTC estimates (Tables: OUTPUT TRTCEstimates CU, OUTPUT TRTCEstimates StatArea).

Field Name	Data Type	Description
SpeciesId	Text	Species
StatArea/CU	Text	Statistical area or conservation unit
CU_Name	Text	Optional - conservation unit name. Not available for statistical area results.
Year	Number	Year
T_Idx_1	Number	Total index stream escapement
ExpFactor1	Number	Expansion Factor 1
AdjSum	Number	Adjusted escapement
ExpFactor2	Number	Expansion Factor 2
ObsE	Number	Observed escapement
ExpFactor3	Number	Expansion Factor 3
TE	Number	Total escapement
CDN Harvest	Number	Canadian harvest
CDN ER	Number	Canadian exploitation rate
TRTC	Number	Total Return to Canada
Total Harvest	Number	Total harvest
Total ER	Number	Total exploitation rate
Total Run	Number	Total run
Q1	Number	Indicator Streams Avg Survey Quality
Q2	Number	Indicator Streams Survey Coverage
Q3	Number	Indicator Streams % of total spawners

Table 6-2. [OUTPUT MeanEscapement] Average stream decadal escapement estimates.

Field Name	Data Type	Description
IndexId	Text	Unique stream by species identifier (combination of SpeciesId and PopId fields).
SpeciesId	Text	Species identification code
StatArea	Text	Statistical area designation derived from the NuSEDS database
CU	Text	Conservation unit code
Decade	Text	Character string indicating the decade
DecadeNo	Integer	Numeric code indicating the dacade
Indicator	Text	Character string indicating whether the stream has been designated as an indicator stream
Total	Integer	Total estimate of escapement for the decade
n	Integer	Number of years within the decade that have available escapement data
ME	Number	Mean decadal escapement

6.2 Returns by Age and Brood Year

Once the TRTC output has been computed for statistical areas and conservation areas (see Section 6.1) the results are then combined with the age tables (see Section 3.3) to produce estimates of total returns by brood year and age (Table 6-3). These results are stored in the [OUTPUT AGE CU] and [OUTPUT AGE Statarea] tables for conservation unit and statistical area results.

Table 6-3. Format for output tables contain the estimates of total return by age and brood year for each CU and statistical area (Tables: [OUTPUT AGE CU], and [OUTPUT AGE Statarea]).

Field Name	Data Type	Description
SpeciesId	Text	Species
StatArea/CU	Text	Statistical area or conservation unit
CU_Name	Text	Optional - Conservation unit name. Not included in statistical area results
BroodYear	Number	Brood year
Escape	Number	Escapement
TR2	Number	Total return at age 2
TR3	Number	Total return at age 3
TR4	Number	Total return at age 4
TR5	Number	Total return at age 5
TR6	Number	Total return at age 6
TR7	Number	Total return at age 7
Total	Number	Total return for all ages
Total ER	Number	Total exploitation rate
Age2	Number	Percent return at age 2
Age3	Number	Percent return at age 3
Age4	Number	Percent return at age 4
Age5	Number	Percent return at age 5
Age6	Number	Percent return at age 6
Age7	Number	Percent return at age 7

6.3 NCC Streams Escapement

As part of the main analysis script (see Appendix B) the yearly stream escapement estimates are updated for the "NCC Streams File" (see Section 5.1). The update extracts all the NuSEDS escapement estimates from the [DATA Escapement] table and combines it with descriptive "meta" data for each stream-species combination (POP_ID) from table [INPUT NCCStreamsFile] and builds a table that can be saved within the System as well as an external MS Excel worksheet (Table 6-4). The source column in Table 6-4 indicates the origin of the information in each data field. LGL refers to work conducted by LGL Limited from 2003-2013. Holtby refers to information provided by Blair Holtby in a dataset he provided to LGL dated 25 October 2011.

Table 6-4. [OUTPUT NCC Streams Escapement] Format for the MS Excel worksheet produced by the NCC streams query.

Excel Column	Field Name	Field Description	Source
A	ID	LGL MS Access database record ID	LGL
B	POP_ID	NuSEDS POP_ID field	Holtby
C	Indicator	Indicator flag (Y=Indicator stream, YY=New Indicator)	LGL
D	SPP	Species	Holtby
E	GFE_ID	NuSEDS population identifier	Holtby
F	SYS_NM	Stream name	Holtby
G	source	Data source (usually NuSEDS)	Holtby
H	popMAP	POP_IDto be used for these data	Holtby
I	EXPN	Proposed expansion factor	Holtby
J	yLAT	Stream location – latitude	Holtby
K	xLONG	Stream location – longitude	Holtby
L	faz_acro	Freshwater Adaptive Zone	Holtby
M	maz_acro	Marine Adaptive Zone	Holtby
N	jaz_acro	Joint Adaptive Zone	Holtby
O	CU_fname	Species+CU name	Holtby
P	CU_facro	Species+CU acro	Holtby
Q	CU_findex	Species+CU index	Holtby
R	CU_name	CU name	Holtby
S	CU_acro	CU acronym	Holtby
T	CU_index	CU index number	Holtby
U	Area	BC geographic area code	Holtby
V	SEP_ENH	Enhancement flag (Y=enhanced)	Holtby
W	IsIndicator	Old Indicator flag (Y=indicator stream)	Holtby
X	IsInFiltEsc	Escapement filter flag (blank in this dataset)	Holtby
Y	nrecs	Old- Number of NuSEDS records for this POPID	Holtby
Z	nins	Old- Number of stream inspections	Holtby
AA	npres	Old- Number of records indicating fish presence	Holtby
AB	nnumest	Old- Number of numerical estimates	Holtby
AC	pins_rec	Old- # inspections records/# NuSEDS records	Holtby
AD	ppres_ins	Old- # presence records/# inspection records	Holtby
AE	pest_pres	Old- # numerical estimates/# presence records	Holtby
AF	fr_timing	Fraser run timing (NA for this dataset)	Holtby
AG	OL_GRP_NM	Group of statistical areas	Holtby
AH	Fraser_mnemonic	Fraser timing group (NA for this dataset)	Holtby
AI	CD_findex-name	Index name (Blank for this dataset)	Holtby
AJ	StatArea	Statistical area code	LGL
AK	Reviewer	Last name of reviewer for indicator streams	LGL
AL	QA	Survey quality code	LGL
AM	Method	Survey method (incomplete)	LGL
AN	Race	Run timing (incomplete)	LGL
AO	WildCode	% Wild code (incomplete)	LGL

AP	WildRigor	% Wild – Data quality code (incomplete)	LGL
AQ	C80s	Number of non-zero escapement years in the 1980s	LGL
AR	C90s	Number of non-zero escapement years in the 1990s	LGL
AS	C00s	Number of non-zero escapement years in the 2000s	LGL
AT	Avg1980-2010	Average of non-zero escapement estimates for 1980-2010	LGL
AU-DC	1950-2010	NuSEDS annual estimates of total escapement	NuSEDS

Source field identifies the source of the information provided in each data field:

- Holtby — Information from the "decoder ring" database prepared by Blair Holtby in October 2011 with meta data for each stream-species combination (i.e., POP_ID).
- LGL — Information from the indicator stream database prepared by LGL Limited through consultations with regional DFO biologists.
- NuSEDS — DFO's April 2018 version of the Salmon Escapement Database System.

7. LITERATURE CITED

Alexander, R., K.K. English, D. Peacock and G. Oliver. 2010. Assessment of the Canadian and Alaskan Sockeye stocks harvested in the northern boundary fisheries using run reconstruction techniques, 2004-08. Report for Pacific Salmon Comm. Northern Boundary Technical Committee.

Challenger, W.O., English, K.K., and Peacock D. 2018. Area 3, 4, 5 Pink and Chum Exploitation Rate Model. Technical report prepared for Pacific Salmon Foundation.

English, K.K. 2013. Extended time-series of catch and escapement estimates for Skeena Sockeye, Pink, Chum, Coho and Chinook salmon conservation units. Prepared for Pacific Salmon Foundation, Vancouver, BC. 19 p.

English, K.K., A. Blakley, T. Mochizuki and D. Robichaud. 2009. Coast-wide review of BC salmon indicator streams and estimating escapement, catch and run size for each salmon conservation unit. Draft report prepared for Fisheries and Oceans, Canada and Ecotrust, Nanaimo, BC. 78 p.

English, K.K. N. Blakley, T. Mochizuki, C. Sliwinski and A. Blakley. 2004a. Fisheries Resource Manuals: North and Central Coast. Prepared for Department of Indian Affairs and Northern Development, Vancouver, BC. 48 p. plus appendices.

English, K.K., W. J. Gazey, D. Peacock and G. Oliver. 2004b. Assessment of the Canadian and Alaskan Sockeye stocks harvested in the northern boundary fisheries using run reconstruction techniques, 1982-2001. Pacific Salmon Comm. Tech. Rep. 13:93 p.

English, K.K., T. Mochizuki and D. Robichaud. 2013a. Review of North and Central Coast salmon indicator streams and estimating escapement, catch and run size for each Salmon Conservation Unit. Report for Pacific Salmon Foundation and Fisheries and Oceans, Canada. 78 p.

English, K.K., C. Noble and A. Blakley. 2013b. Skeena sockeye in-river run reconstruction analysis model and analysis results for 1982-2009. Report for Pacific Salmon Foundation. 42 p.

English, K.K., D. Peacock and B. Spilsted. 2006. North and Central Coast core stock assessment program for salmon. Prepared by LGL Limited for the Pacific Salmon Foundation and Fisheries and Oceans Canada. 78 p.

Gazey, W.J. and K.K. English. 2000. Assessment of sockeye and pink salmon stocks in the northern boundary area using run reconstruction techniques, 1982-95. Can. Tech. Rep. Fish. Aquat. Sci. 2320:132 p.

Holtby, L.B. and K.A. Ciruna. 2007. Conservation units for Pacific salmon under the Wild Salmon Policy. Canadian Science Advisory Secretariat, Research Document 2007/070.

Korman, J. and K.K. English. 2013. Benchmark analysis for Pacific Salmon Conservation Units in the Skeena Watershed. Prepared for Pacific Salmon Foundation, Vancouver, BC. 88 p.

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: https://www.R-project.org/

8. APPENDICES

APPENDIX A. SYSTEM SETUP

The 2018 version of the NCC Salmon Database System uses the R computing environment (R Core Team 2018) to interface and update the Microsoft Access database. R is a programming language and free software environment for statistical computing and graphics that is supported by the R Foundation for Statistical Computing. The R language is widely used among statisticians and scientists and can be downloaded and installed for free.

A basic understanding of computer programming and the R Computing Environment is only required to update the database, while only Microsoft Access is required to view an updated database. For convenience, the main analysis script also produces a series of Microsoft Excel files with main output data for easier dissemination.

R interfaces with the NCC Salmon Database (a Microsoft Access database) through the Open DataBase Connectivity (ODBC) standard, which requires the Microsoft "ACE" ODBC driver (see below for further details). Because all interfacing with the NCC Salmon Database is through the ODBC standard, the NCC Salmon Database can be switched to another SQL database backend, such as PostgreSQL, an open-source alternative, with only a small modification to the existing codebase.

Software requirements to view an updated NCC Salmon Database are listed separately from the requirements to update the NCC Salmon Database as the latter requires installation and setup of R.

Software required to view the NCC Salmon Database:

- Microsoft Windows 10;
- Microsoft Access 2010 or newer; and
- The NCC Salmon Database file (*NCC Salmon Database v2.accdb*).

Additional software required to update the NCC Salmon Database:

- Microsoft Access Driver (*.mdb, *.accdb):
    - The newer "ACE" ODBC driver. It is not included with Windows, but it is normally included as part of a Microsoft Office install. It is also available as a free stand-alone "redistributable" installer for machines without Microsoft Office;
    - Installer: https://www.microsoft.com/en-us/download/confirmation.aspx?id=13255;
- the R computing environment:
    - Installer: https://cran.r-project.org/;

  - Pre-requisite R packages must also be installed in the R environment, which is handled by the main analysis script, but can also be done manually (see *R, R packages and RStudio* section below for further details);

- The main R analysis script (i.e., *nccsdb-analysis.R*) containing all the required commands and the data import settings (i.e., *import-settings.R*);
- The custom NCCSDB R-package containing specific routines to import data and carry out required computations and data manipulations:
  - This R-package will be distributed within the NCC Salmon Database, and can be found in a directory called /NCCSDB;
- *Optional* – The RStudio Desktop application:
  - Integrated Development Environment (IDE) for common actions and tasks conducted in R;
  - Installer: https://www.rstudio.com/products/rstudio/download.

R packages, and RStudio:

If you are updating the NCC Salmon Database, once the base R computing environment has been installed, prerequisite R-packages (Table A-1) must also be installed within the R environment before attempting to update the NCC Salmon Database. R-packages extend the base functionality of R and used by the custom NCCSDB package. Packages can be installed either by text commands in the R console, or through the graphical user interface (GUI) in the R desktop application or the RStudio IDE.

For the sake of convenience, the main analysis script (i.e., *nccsdb-analysis.R*) contains three lines of code to check for missing packages and then attempts to download and install any that are missing:

```
package.list <- c('devtools', 'openxlsx', 'plyr', 'reshape2', 'RODBC', 'roxygen2', 'stringr')
new.packages <- package.list[!(package.list %in% installed.packages()[,"Package"])]
if(length(new.packages) > 0) install.packages(new.packages)
```

If the above command fails to install the R packages, the missing R packages can be manually installed via the R desktop application or the RStudio IDE.

Once the R packages have been installed the main analysis script will then load the custom NCCSDB package, which will ensure the required packages have been properly installed. The custom NCCSDB package should be included with the files distributed with the NCC Salmon Database.

Selecting the Correct Executable Version of R

There are two available versions of the ODBC driver for Microsoft access: a 32-bit and 64-bit version. In order for R to connect the NCC Salmon Database through the ODBC driver (via the RODBC R-package) the technology stack from R onwards must be in the same bit version (Figure A-1). If the wrong R executable version is used the RODBC package will not be able to correctly connect to the Microsoft Access ODBC driver, resulting in an error message indicating: "no default driver specified." Microsoft currently provides two different installer files for 32-bit

technology stack (i.e., *AccessDatabaseEngine.exe*) and for a 64-bit technology stack (i.e., *AccessDatabaseEngine_X64.exe*). The main analysis script relies on the `odbcConnectAccess2007` function in the RODBC R-package, which requires the 32-bit version of the Microsoft Access Driver. 64-bit technology stack should be possible, but has not been tested so it is not currently a recommended as a connection method.

Once the ODBC correct driver is installed the R computing environment must also be run in the same bit version. When the base R installer is run, both a 32-bit and 64-bit version are installed if you are running the 64-bit version of Windows operating system. If you are using the basic R GUI desktop application included with the base R installation file, then the separate 32-bit (e.g., *R i386 3.5.1.exe*) should be used.

If you are using RStudio IDE (recommended) then you will need to update the RStudio global setting to use the 32-bit version of R (Figure A-2). The settings dialog can be found by under the under the *Tools* > *Global Options…* menu and should be the first option presented. Once the 32-bit version of R is selected, RStudio will need to be re-launched so that the correct version of R is loaded at the time of application startup.

Table A-1. List of prerequisite R-packages required by the custom NCCSDB R-package used to carry out all NCC Salmon Database manipulations.

R-Package	Purpose	Further Information
`devtools`	Development tools used for preparing and loading the custom NCCSDB package used for all calculations and manipulations.	https://cran.r-project.org/web/packages/devtools/
`openxlsx`	Tools for reading and writing to Microsoft Excel files.	https://cran.r-project.org/web/packages/openxlsx/
`plyr`	Tools for splitting, applying, summarizing, and combining data	https://cran.r-project.org/web/packages/plyr/
`reshape2`	Tools for Flexibly restructure and aggregate data	https://cran.r-project.org/web/packages/reshape2/
`RODBC`	An R implementation of ODBC (Open DataBase Connectivity), which allows access to a variety of database systems via a common interface.	https://cran.r-project.org/web/packages/RODBC/
`roxygen2`	In-line documentation for R packages, used to generate help files for the custom NCCSDB package,	https://cran.r-project.org/web/packages/roxygen2/
`stringr`	A consistent, simple and easy to use set of wrapper functions for string manipulations.	https://cran.r-project.org/web/packages/stringr/

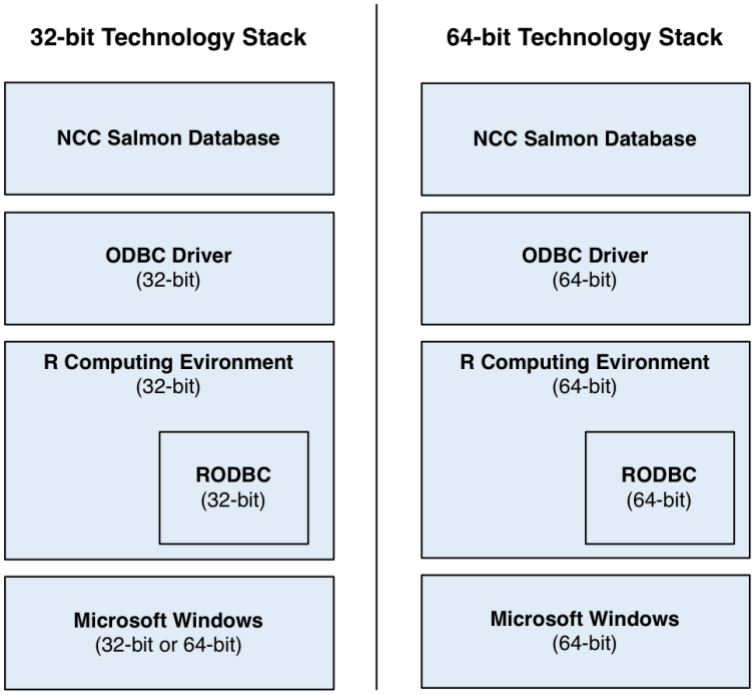

Figure A-1. Possible technology stacks for connecting R to the NCC Salmon Database (currently only the 32-bit stack has been tested and verified).

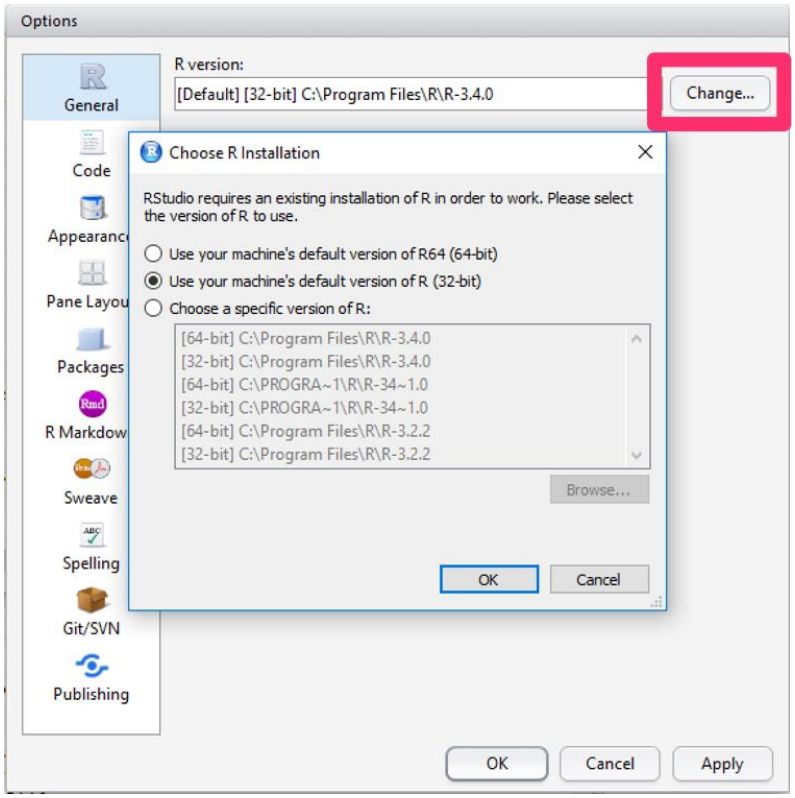

Figure A-2. RStudio settings dialog box which allows the 32-bit or 64-bit version of R to be selected.

Troubleshooting

Almost all troubleshooting issues associated with updating the NCC Salmon Database can be traced back to how R interacts with the ODBC driver. Errors can be the result of an incorrect file path, leaving the NCC Salmon Database open in Microsoft Access (i.e., this can prevent lock tables from being updated, resulting in errors about tables not being available), or a mismatch between the R bit version and the ODBC driver bit version (Figure A-2).

If database connection errors are indicated, the following steps, in order of priority, should be taken to ensure attempt to resolve the connection issue:

1. Ensure that you have specified the correct file path the NCC Salmon Database;
    a. Sometime this requires specifying the full file path instead of a relative file path;
    b. Try using the built-in base commands `file.path,` and `file.exists` to verify your full file path;
2. Ensure that the NCC Salmon Database is not currently open in Microsoft Access (i.e., open tables can lock that table from being updated);
3. Re-install the 32-bit version of the Microsoft Access Drivers (i.e., installer: *AccessDatabaseEngine.exe);*
    a. Updates to Microsoft Windows can re-install the 64-bit ODBC driver;
4. Ensure that RStudio is using the 32-bit version of R (see Figure A-2);
5. Ensure that Windows ODBC settings has a 32-bit entry for Microsoft Access .accdb files (Figure A-3);
    a. *Optional* - If an entry does not exist in the 32-bit registry, then a new data source connection type for .accdb files (Figure A-4) by clicking on the “Add…” button in the 32-bit ODBC Data Source Administrator dialog (Figure A-3).

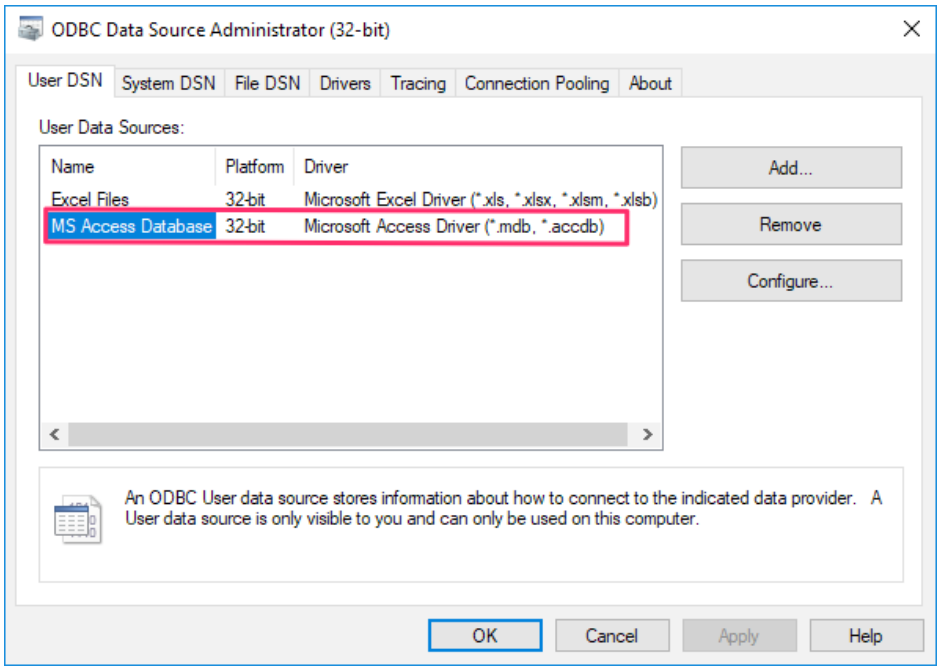

Figure A-3. Under the Windows settings check the 32-bit ODBC Data Source Administrator registry has the appropriate entry for Microsoft Access Driver for .accdb compatibility.

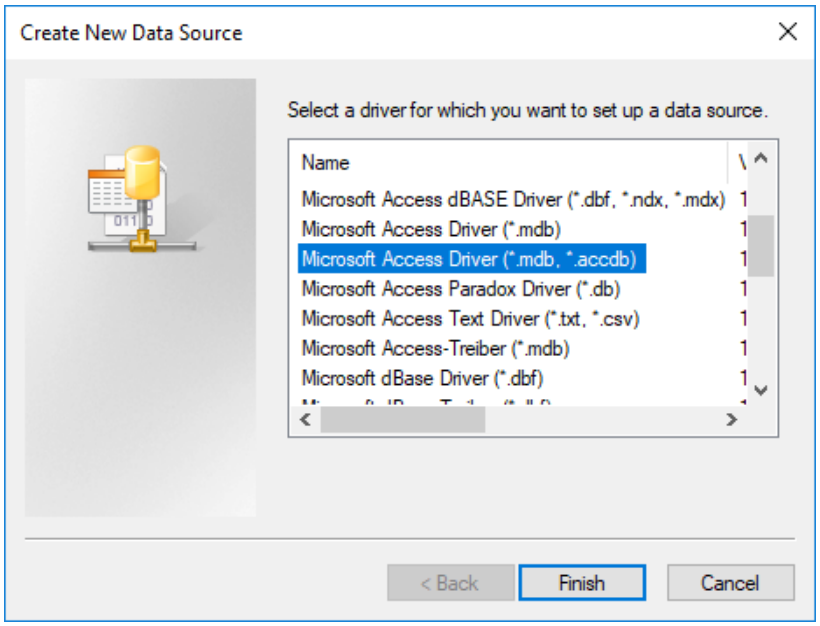

Figure A-4. Optional – if a Microsoft Access ODBC Driver for .accdb files is not present, add an entry using the "Create New Data Source" dialog.

APPENDIX B. RUNNING THE ANALYSIS SCRIPT

The main script (i.e., *nccsdb-analysis.R*) contains all the commands to import the primary data (see Section 3), import data from external analyses, including data processing (see Section 4), import tables containing analysis settings (see Section 5), and all the steps to compute all the final output results (see Section 6). Results are saved back to various OUTPUT tables within database, as well as to Microsoft Excel workbooks for convenience (see Section 6). Due to the large number of settings required to import data from a variety of sources, all of these settings are kept in data a separate setting file (i.e., *import-settings.R*). This allows all the settings associated with a particular analysis year to be easily archived. Both scripts have been formatted using R Studio section headings, which allows for quick navigation (Figure B-1).

Once the import settings have been correctly specified the entire analysis script can be run in one step by clicking on the "Source" button in R studio (Figure B-1a). This runs all of the commands in the script in succession. Note that the main analysis script (i.e., *nccsdb-analysis.R*) includes a small section at the top of the file called "User Settings" for general user settings such as file paths and analysis years (Figure B-1a). It is not recommended that the script edited below this section.

Before running the main analysis script ensure the database is not currently open in Microsoft Access. Keeping a table open in Microsoft access will lock it from changes, resulting in an error (i.e., a message about a table already existing will be displayed in the RStudio console).

Importing External Analyses Data

Routines to import primary data sources (see Section 3) are straight forward and generally do not require user interventions, the same is not true of external analysis data. Some species-specific results are only read in from a portion of an Excel Worksheet, with the portion read depending on range of analysis years. As such, the worksheet and cell range each external analysis data source have been defined exactly in *import-settings.R* and will need to be re-defined (or checked) for any updates to the external analysis Excel Workbooks.

To facilitate this task, the `view.data` variable in "User Settings" section of the main analysis script (i.e., *nccsdb-analysis.R*) provides the option to open and review each external analysis data before committing to importing the data into the NCC Salmon Database. When set to `TRUE`, a separate data viewer window is launched for each external analysis data source (Figure B-2 and Figure B-3). At this point, the main script will be paused allowing the user to inspect each data source. Once satisfied that the external analysis data is correct (e.g., all required years are present), the main analysis script can be re-initiated by navigating to the RStudio console window and indicating that you wish to proceed (Figure B-3). Note that setting `view.data` variable to `FALSE` (see Figure B-2) will skip this step allowing the analysis script to fully execute without any user interventions.

Finally, the codebase was designed to be verbose. As the above steps are carried out a stream of updates and status results will be displayed in the RStudio console window (Figure B-3). Ensure that these are reviewed before relying on the NCC Salmon Database output.

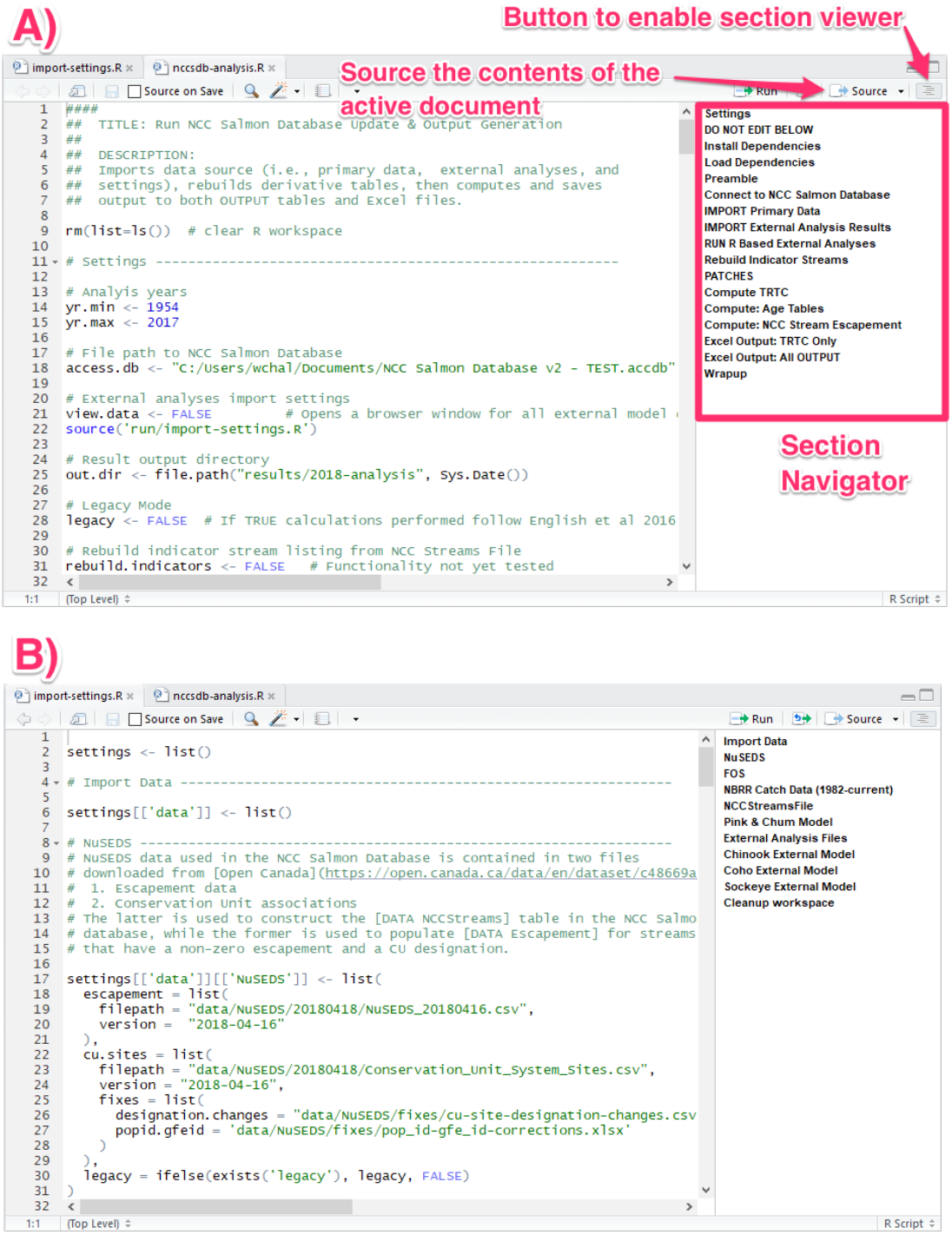

Figure B-1. The main analysis script (A) and the import setting script (B) have been formatted for easy navigation in R Studio.

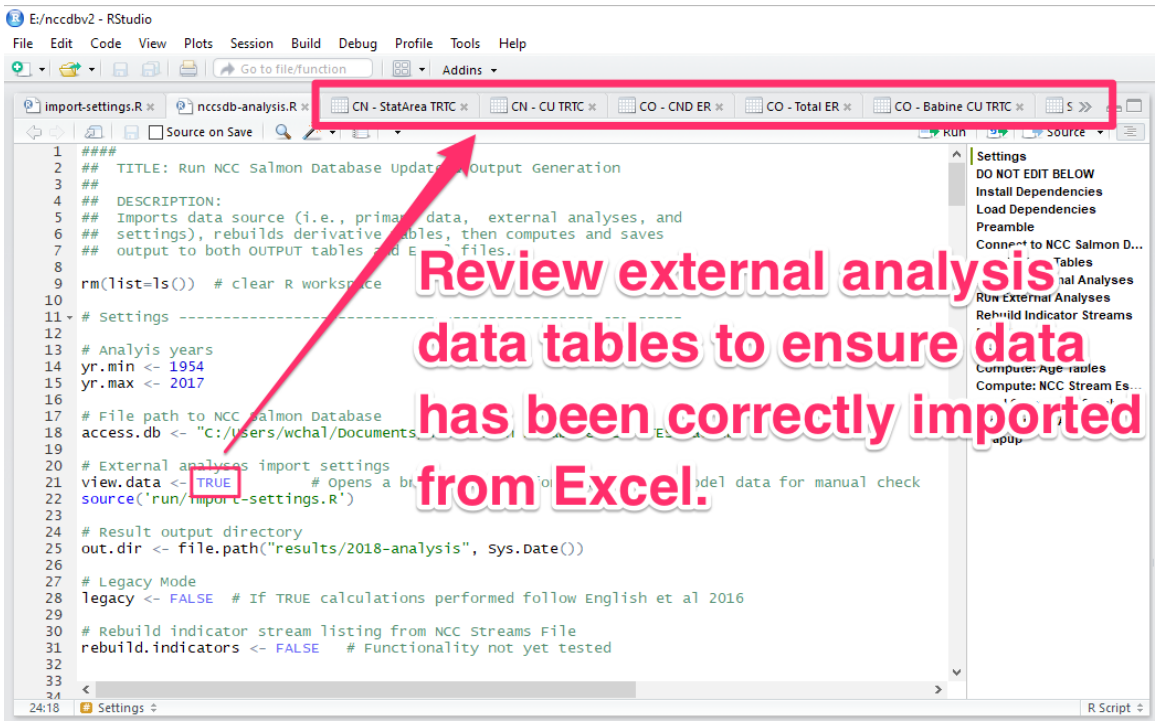

Figure B-2. **Enabling the `view.data` option will open separate viewer windows for each of the external analysis results to be imported so that import settings can be confirmed.**

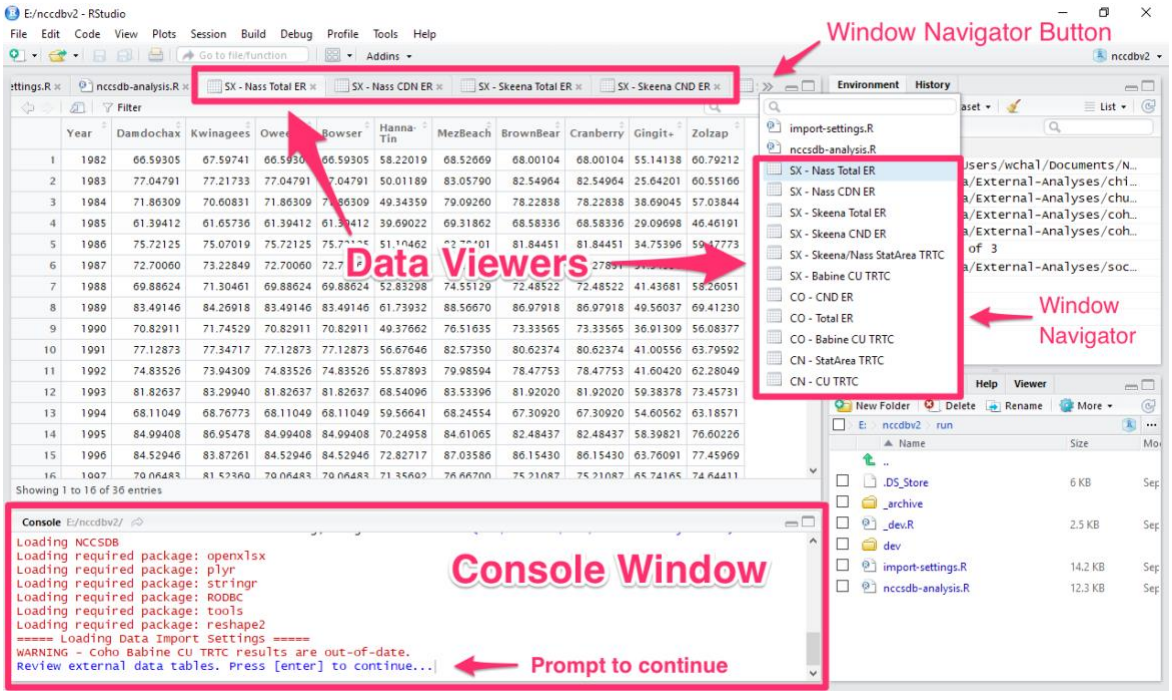

Figure B-3. **Script execution pauses allowing the user to review external analysis data in the data viewer windows, when satisfied the script can be continued by navigating to the console window and hitting the "enter" key.**

APPENDIX C. NCCSDB R-PACKAGE

The main analysis script (see Appendix B) can be viewed as containing the high-level business logic associated with updating data within the System and then compiling the final output data. Most of the actual tasks are performed by routines (i.e., functions) contained in a custom R-package developed for the project (Table C-1). This package is referred to within this document as the NCCSDB custom R-package.

Table C-1. List of functions included in the NCCSDB R-package along with a description of the functions purpose and any associated report sections.

Function	Class	Report Section	Description
BlankTRTC	TRTC	Section 6.1	Creates a new TRTC form without values
BlankWorkSheet	Utility		Creates a new Excel workbook worksheet or if the worksheet is already present it clears out all values.
CalculateStreamEscapement	Data Import	Section 3.1 & Appendix D	Computes NuSEDS stream escapement
ClearDataTables	Database Utility		Truncates (i.e., deletes all records) in DATA tables. Useful to ensure all records in the data base are from intended data sources
ClearInputTables	Database Utility		Truncates (i.e., deletes all records) in INPUT tables. Useful to ensure all records in the data base are from intended data sources
ClearOutputTables	Database Utility		Truncates (i.e., deletes all records) in INPUT tables. Useful to ensure all result records are from the curren analysis.
CompareSchema	Database Utility		Utility function used to compare schema between two Microsoft Access data bases. Useful for determining structural changes that may have occurred between database versions.
CompareTRTC	TRTC		Utility function that compares and summarizes differences between TRTC tables. This includes determine if any new or missing records exist in addition to changes in values for common records. Used to compare how TRTC output results have changed with input data revisions or analysis parameter changes.
ComputeEscapement	TRTC	Section 6.1	Computes escapement by geographical area (e.g., statistical area or conservation unit).
ConservationUnitER	TRTC	Section 6.1	Computes all of the conservation unit exploitation rates by calling the appropriate species specific sexploitation rate functions.
ConservationUnitERChinook	TRTC	Section 6.1	Computes conservation unit exploitation rates for Chinook salmon.
ConservationUnitERChum	TRTC	Section 6.1	Computes conservation unit exploitation rates for Chum salmon.
ConservationUnitERCoho	TRTC	Section 6.1	Computes conservation unit exploitation rates for Coho salmon.
ConservationUnitERPink	TRTC	Section 6.1	Computes conservation unit exploitation rates for Pink salmon.
ConservationUnitERSockeye	TRTC	Section 6.1	Computes conservation unit exploitation rates for Sockeye salmon.
ConservationUnitTRTC	TRTC	Section 6.1	Computes the TRTC output results for conservation units.

Function	Class	Report Section	Description
Convert2StatArea	Data Import	Section 3.1 & Appendix D	Utility function that converts NuSEDS AREA field values to the NCC Salmon database StatArea designation (see Table 3-2).
ConvertPOPID	Data Import	Section 5.1	Converts POP_ID values between the NuSEDS implementation and the Blair Holtby revision used to distinguish between odd and even year Pink salmon, provided to LGL on October 25, 2011.
ConvertSpeciesCode	Data Import	Section 3.1 & Appendix E	Utility function designed to replace ConvertSpeciesQualified()
ConvertSpeciesQualified	Data Import	Section 3.1 & Appendix E	Utility function that converts between species designations used across the different data bases (see Table 3-3).
ExcelColNum	Data Import	Section 4 & Appendix D	Utility function that converts Excel's A1 style column reference style into an integer index for compatibility with other R routines.
ExpansionFactor2	TRTC	Section 6.1	Computes Expansion Factor 2 values for statistical areas or conservation units
ExpansionFactor3	TRTC	Section 6.1	Computes Expansion Factor 3 values for statistical areas or conservation units
ExpFactor1RefDecade	TRTC	Section 6.1 & 5.6	Determines reference periods (i.e., reference decades or longer periods of time if no suitable decades are found).
GetColNumbers	Data Import	Section 4 & Appendix D	Utility function that converts ranges or series of Excel's A1 style column reference style into a range or series of integer index for compatibility with other R routines.
ImportFOS	Data Import	Section 3.2	Reads in, validates, and then imports FOS data.
ImportHistPinkChumData	Data Import	Section 3.4	Reads in, validates, and then imports Historical Pink & Chum Model data into the appropriate tables.
ImportNBRRCatchEffort	Data Import	Section 3.4	Imports current BC and Alaskan catch and effort data from the Sockeye Northern Boundary Run Reconstruction database.
ImportNCCStreamsFile	Data Import	Section 5.1	Imports the "NCC Stream File" after stripping off escapement values.
ImportNuSEDS	Data Import	Section 3.1	Reads in, validates, and then imports NuSEDS escapement data.
IndicatorEscapement	TRTC	Section 6.1	Computes the indicator escapement (i.e., the T_Idx_E field in the TRTC output).
InsertAccessRecords	Database Utility		Inserts data records into Access, handles spaces in field names and runs some checks (e.g., destination field names match input data) before inserting records.
LoadExcelTable	Data Import	Section 4	Loads data from a specified Excel worksheet and cell range, also executes a clean-up function if specified.
NCCStreamEscapement	Output	Section 6.3	Appends yearly stream escapement to the "NCC Streams File"
NuSEDSStreamSearch	Utility		Utility function that attempts to find a stream name by partial string match of various fields in the NuSEDS database.

Function	Class	Report Section	Description
PinkChumModel	External Analysis	Section 4.1.2	Pink & Chum external analysis model used to produce Area 03-05 exploitation rates for Pink and Chum salmon.
PinkChumModelDiagnosticOutput	External Analysis	Section 4.1.2	Provides data summaries and diagnostics used to assess the Pink & Chum Model.
PrepareBabineCOTRTC	Data Import	Section 4.2	Prepares the Babine Coho TRTC external analysis results for insertion into the database.
PrepareChinookTRTC	Data Import	Section 4.2	Prepares the Chinook statistical area TRTC external analysis results for insertion into the database.
PrepareChumPinkER	Data Import	Section 4.1.2	Prepares Pink and Chum exploitation rate estimates for database insertion.
PrepareCohoER	Data Import	Section 4.1.3	Prepares Coho exploitation rate estimates for database insertion.
PrepareSockeyeBabineTRTC	Data Import	Section 4.2	Prepares the Babine Sockeye TRTC external analysis results for database insertion.
PrepareSockeyeER	Data Import	Section 4.1.1	Prepares Coho exploitation rate estimates for database insertion
PrepareSockeyeTRTC	Data Import	Section 4.2	Prepares the Sockeye statistical area TRTC external analysis results for database insertion.
ProcessChinookExtModData	Data Import	Section 4.3	Post-processing of Chinook external analysis data.
ProcessCohoExtModData	Data Import	Section 4.3	Post-processing of Coho external analysis data.
ProcessPinkChumExtModData	Data Import	Section 4.3	Post-processing of Pink and Chum external analysis data.
ProcessSockeyeExtModData	Data Import	Section 4.3	Post-processing of Sockeye external analysis data.
RebuildIndicatorStreams	Data Import	Section 5.1	Rebuilds the list of indicator streams based on values in the “NCC Streams File”
ReturnsByAge	Brood Tables	Section 6.2	Produces the brood year returns by age results.
StatAreaERAvg	TRTC	Section 6.1	Computes the average exploitation rate across a set of statistical areas.
StatAreaTRTC	TRTC	Section 6.1	Computes the TRTC output for statistical areas
SummarizeERSource	TRTC	Section 6.1	Summarizes the source model for exploitation rate estimates
SurveyRatingQ1	TRTC	Section 6.1	Retrieves the survey quality rating for each stream
TotalCatch	TRTC	Section 6.1	Computes the species-specific yearly total catch for statistical areas from the FOS catch data
UpdateExternalAnlayses	Data Import	Section 4.1 & 4.2	General function used to update species-specific external analysis results based on the specified settings.
UpdateExtModData	Data Import	Section 4.3	Utility function that imports the external analysis results tables after running a series of checks. Also confirms that imported data matches the original data.
UpdateLookupCU	Data Import	Section 5.4	Updates the [LOOKUP CU] table based on a variety of data sources.

Function	Class	Report Section	Description
UpdateNCCStreams	Data Import	Section 3.1 & Appendix E	Updates the [DATA NCCStreams] table based on the Conservation Unit System Site file and NuSEDS escapement data.
UpdateNuSEDSEscapement	Data Import	Section 3.1 & Appendix D	Updates the [DATA Escapement] for NCC streams table based on the supplied version of NuSEDS.
UpdateTable	Utility		Utility function used to selectively updated values in a data frame.
WriteValidation	Utility		Utility function used to write out model results to a validation file as part of unit testing.
YearRun	Utility		Utility function used to summarize contiguous runs of years (e.g., 1,2,3,4,7,8,9 would be summarized as 1-4,7-9).

APPENDIX D. EXTRACTING ESCAPEMENT DATA FROM THE NUSEDS DATABASE

Stream escapement data is derived directly from the NuSEDS database. Typically, the MAX_ESTIMATE field will contain the maximum final estimate, but this field may not always be available. As such, escapement is determined based on data available in the following NuSEDS fields:

- NATURAL_ADULT_SPAWNERS;
- NATURAL_JACK_SPAWNERS;
- NATURAL_SPAWNERS_TOTAL;
- ADULT_BROODSTOCK_REMOVALS;
- JACK_BROODSTOCK_REMOVALS;
- TOTAL_BROODSTOCK_REMOVALS;
- OTHER_REMOVALS;
- TOTAL_RETURN_TO_RIVER; and
- UNSPECIFIED_RETURNS.

The available data for each stream can vary greatly and as such the algorithm is used to determine the escapement for each year. The `CalculateStreamEscapement()` routine in the NCCSDB package handles this task (Appendix C) . It is also expected that data extracted from the NuSEDS database will be provided in the format described in Table D 1. If the data format differs from what is outline Table D 1 then the routine may fail.

If the UNSPECIFIED_RETURNS field is available this is used as the escapement estimate, otherwise escapement is computed by the following steps:

1. Determine the total number of spawners by using :
    a. Use NATURAL_ADULT_SPAWNERS if available, otherwise
    b. Use NATURAL_SPAWNERS_TOTAL
2. Determine the broodstock removals:
    a. Use ADULT_BROODSTOCK_REMOVALS if available, otherwise
    b. Use TOTAL_BROODSTOCK_REMOVALS
3. Determine if other adult removals are present (i.e., OTHER_ADULT_REMOVALS)
4. Calculate escapement as Spawners (1) + Broodstock (2) + Other Removals (3)

Finally, if this does not produce an escapement estimate the TOT_ADULT_RET_RIVER field was used, then finally the MAX_ESTIMATE field was used as a final fall back. Currently, within NCC based streams we found no discrepancy between values present in the MAX_ESTIMATE field and those computed via the outlined steps.

The yearly escapement estimates are saved to the “Return” field within [DATA Escapement] table (Table 3-4) along with the species designation and year and other internal fields specified

in Table 3-2. The unique NuSEDS record identifier (i.e., ACT_ID) is also included as the “Id” field, to allow escapement records to be traced back to the source.

Table D 1. Format for data received from the NuSEDS database.

Field Name	Type	Description	Destination Table	Destination Field
AREA	Text	This is the subdistrict. In most cases subdistricts are the same as statistical areas. They mainly differ for streams that eventually drain into the Fraser and for large areas that have been split up and thus have a/b/c... designations. E.g. Statistical area 03 has two subdistricts 3A and 3B.	[DATA Escapement] & [DATA NCCStreams]	StatArea
WATERBODY	Text	This is the name of the waterbody or portion of a waterbody that bounds the population as shown on any given SEN.		
GAZETTED_NAME	Text	Provincially recognized name for the waterbody.		
LOCAL_NAME_1	Text	Commonly known name for the waterbody.		
LOCAL_NAME_2	Text	Second most commonly known name for the waterbody.		
WATERSHED_CDE	Text	45 digit hierarchical provincial code unique to the waterbody and its watershed.		
WATERBODY_ID	Text	This is a combination of 5 digits that uniquely identify a GIS polygon and four characters that uniquely describe a provincial watershed group.		
RAB_CDE	Text	Discontinued Resource Analysis Branch Code unique to each waterbody.		
POPULATION	Text	Default naming originates from previous databases as a concatenation of stream name, subdistrict, species and run type. This is the most important piece of data that all the other SEN data fields refers to.		
RUN_TYPE	Text	Run_Type indicates the run timing for different runs within the same season. In some cases, the runs may be well documented enough to label them as something like "Spring" vs. "Summer" or "Early vs. "Late". But in other cases there is no documentation other than to indicate that there are distinct runs within a season. In these cases, we used the numeric labelling approach for historic data in order to avoid adding unintentional (and potentially inaccurate) detail.		
SPECIES	Text	Species of Fish.	[DATA Escapement] & [DATA NCCStreams]	SpeciesId & IndexId
ANALYSIS_YR	Integer	This is the year that the estimate is for. Surveys may have continued into the following calendar year.	[DATA Escapement]	Year

Field Name	Type	Description	Destination Table	Destination Field
START_DTT	Text	This is the time stream inspections began e.g. 2000-10-15 means that the first inspection for this season's estimate started on October 15 2000.		
END_DTT	Text	This is the time stream inspections ended e.g. 2000-11-15 means that the last inspection for this season's estimate started on November 15 2000.		
ESCAPEMENT_ANALYST	Text	Person responsible for estimate(s) on this SEN.		
ACCURACY	Text	This is the ability of a measurement to match the actual value of the quantity being measured. Some historical estimates that were imported had reliability data originating from SEDS that may appear here.		
PRECISION	Text	This is the ability of a measurement to be consistently reproduced, or put another way, the number of significant digits to which a value has been reliably measured.		
INDEX_YN	Text	This indicates whether the estimates are for a portion of the population. This is usually due by purposely limiting enumeration to a portion of the spawning habitat or a portion of the duration of the run.		
RELIABILITY	Text	This field was added for the inclusion of historical data from an external source. It is the level of reliability that the person placed in their annual estimate of adults. Since this was only recorded for some historical BC16s it will not be visible in all cases. Values are low, medium low, medium, medium high and high.		
ESTIMATE_STAGE	Text	Preliminary SENs are the first drafts of summary estimate documents. Source data may be incomplete and their accuracy has not been verified. Significant changes from Preliminary estimates are probable. Near Final SENs are based on data that have been verified for completeness and accuracy. Further analysis may take place. Final data verification and analysis have not been completed. Minor changes in Near Final estimates are possible. Final SENs are released after all data have been incorporated into the analyses and all verification steps have been completed. Changes are not anticipated.		
ESTIMATE_CLASSIFICATION	Text	This categorizes estimates based on their levels of accuracy and precision (Type-1 are the most accurate, Type-6 the least accurate).		

Field Name	Type	Description	Destination Table	Destination Field
		There are three other classifications that belong to SENs whose source data were migrated from the regional MSAccess SILBC16 database (definitions extracted from that user manual). RELATIVE: CONSTANT MULTI-YEAR METHODS and RELATIVE: VARYING MULTI-YEAR METHODS: "This is the case with survey methods restricted to a fraction of the spawning habitat and/or a fraction of the spawning period. There are various types of relative abundance estimates depending on the survey method, the level of standardization of the methods, and the sampling effort. For our purpose we have retained one type based on between-year consistency of the method where there are two levels." NO SURVEY THIS YEAR: "stream was not inspected for that species this year"		
NO_INSPECTIONS_USED	Integer	This is the number of stream inspection logs that are linked to the SEN or were used in the analysis. E.g. 10 stream inspections and a fixed site survey may have been done in the season, but only 7 stream inspections and the fence counts will be used to produce the annual estimate(s), and only these are linked to the SEN.		
ESTIMATE_METHOD	Text	There are several standard methods to chose from. Addition/Subtraction - simple addition or subtraction to provide an estimate. Should be used in conjunction with activity types Adjustment/Calibration and Summary observations. E.g. a population aggregate, the sum of two or more populations, would require the linking of two or more SENs and straight summation of the estimates. Multiplication/Division - simple multiplication or division to summary estimates. This method should be used in conjunction with activity type Adjustment/Calibration. E.g. E.g. An annual estimate that was arrived at by Peak Live Plus Dead analysis can be adjusted by some factor to make it equivalent to a Time Series estimate that uses AUC calculations. Area Under the Curve - Combining a series of point estimates for abundance to create an estimate for the annual abundance. This is done by determining the total area under a curve of abundance by time then dividing by the survey life (the average length of time that an individual is available to be observed alive i.e. is still within the survey area and is not dead).		

Field Name	Type	Description	Destination Table	Destination Field
		Peak Live Plus Dead - Examine point estimates for abundance, determine the survey when the maximum live count observed; sum the live and dead counts for that survey to create the annual estimate. Peak Live Plus Cumulative Dead - Examine point estimates for abundance, determine the survey when the maximum live count observed. Sum the live count for that survey with the cumulative total of the dead counts prior to and including that survey to create the annual estimate. Fixed Site Census - Combining one or more raw observations into a single estimate (e.g. add all daily fence observation SIL to create a single annual estimate). Mark and Recapture - Petersen - Use capture and re-capture SIL data to determine an abundance estimate with the Petersen calculation. Mark and Recapture - Jolly-Seber - Use capture and re-capture SIL data to determine an abundance estimate with the Jolly-Seber calculation. Redd Count - Using counts of redds from SILs and multiplied by a factor such as 2. Lake Expansion - expanding the dead recoveries by the recovery effort Cumulative New - N/A		
STREAM_ARRIVAL_DT_FROM	Text	This is the start date when the fish first arrive in the water body described on the SEN Details page. Note that the spawn run timings are paired so that Arrival, Start, Peak, and End each have beginning and end date values to represent a date range. The following definition applies to SENs whose source data were migrated from the regional MSAccess SILBC16 database applying mainly to areas 11 to 27, and Fraser chinook/coho, all from 1995 to 2001: "is defined as the month and days (period) that 5% of the fish arrived in the stream. If the number at peak spawning is known, you can identify any of your counts that correspond to 5% of that value. If so the date you made this observation will correspond to the arrival date."		
STREAM_ARRIVAL_DT_TO	Text	This is the end date of arriving fish to the water body described on the SEN Details page.		

Field Name	Type	Description	Destination Table	Destination Field
START_SPAWN_DT_FROM	Text	This is the spawning start date for a population for the current season where start means fish are beginning to pair on the spawning grounds, schools of fish may be holding (pools or mouth) and there are very few, if any, carcasses or redds. Fish are generally in the lower sections of the normal spawning area and may be bright with no fungus and have no white-coloured, eroded fins. The following definition applies to SENs whose source data were migrated from the regional MSAccess SILBC16 database applying mainly to areas 11 to 27, and Fraser chinook/coho, all from 1995 to 2001: "month and days (period) when fish are paired and redds are observed. "		
START_SPAWN_DT_TO	Text	This is the end date of the start of spawning period. See above for the definition of this run timing period.		
PEAK_SPAWN_DT_FROM	Text	This field records the Spawning Peak date for a population and a given season. Peak means the majority of the fish present are paired and actively spawning with few fish holding. The fish may have fungus or white-coloured, eroded fins. A significant proportion of the spawning grounds should have evidence of redds and the fish should generally be distributed throughout the spawning area. The following definition applies to SENs whose source data were migrated from the regional MSAccess SILBC16 database applying mainly to areas 11 to 27, and Fraser chinook/coho, all from 1995 to 2001: "month and days (period) when the number of fish spawning reached its maximum. "		
PEAK_SPAWN_DT_TO	Text	This is the end date of the peak of spawning period. See above for the definition of this run timing period.		
END_SPAWN_DT_FROM	Text	This field records the Spawning End date for a population and a given season. End means very few fish are on the spawning grounds, few unspawned fish are holding and there are lots of carcasses. The remaining fish will likely occupy the upper reaches of the spawning area. The following definition applies to SENs whose source data were migrated from the regional MSAccess SILBC16 database applying mainly to areas 11 to 27, and Fraser chinook/coho, all from 1995 to 2001: "month and days (period) when virtually all fish have spawned in the stream."		
END_SPAWN_DT_TO	Text	This is the end date of the end of spawning period. See above for the definition of this run timing period.		

Field Name	Type	Description	Destination Table	Destination Field
ADULT_PRESENCE	Text	Values are present if adults were observed, none observed if no adults were observed during the stream inspections, not inspected if adults were not looked for, unknown if it is not known whether adults were observed during inspections or not.		
JACK_PRESENCE	Text	Values are present if jacks were observed, none observed if no jacks were observed during the stream inspections, not inspected if jacks were not looked for, unknown if it is not known whether jacks were observed during inspections or not.		
MAX_ESTIMATE	Integer	Is the maximum estimated number taken from: NATURAL_ADULT_SPAWNERS, NATURAL_JACK_SPAWNERS, NATURAL_SPAWNERS_TOTAL, ADULT_BROODSTOCK_REMOVALS, JACK_BROODSTOCK_REMOVALS, TOTAL_BROODSTOCK_REMOVALS, OTHER_REMOVALS, TOTAL_RETURN_TO_RIVER, UNSPECIFIED_RETURNS.	[Data Escapement]	Return
NATURAL_ADULT_SPAWNERS	Integer	All salmon that have reached maturity, excluding jacks (jacks are salmon that have matured at an early age).		
NATURAL_JACK_SPAWNERS	Integer	These are fish that have matured at an early age and are considered precocious. They are usually distinguished from adults by their small size.		
NATURAL_SPAWNERS_TOTAL	Integer	This is the sum of adult and jack natural spawners		
ADULT_BROODSTOCK_REMOVALS	Integer	All salmon that have reached maturity, excluding jacks (jacks are salmon that have matured at an early age) that have been removed from the natural environment for artificially pairing and incubation of progeny in an artificial environment for at least some portion of the incubation period. Eg. hatchery broodstock.		
JACK_BROODSTOCK_REMOVALS	Integer	these are fish that have matured at an early age and are considered precocious that have been removed from the natural environment for artificially pairing and incubation of progeny in an artificial environment for at least some portion of the incubation period. Eg. hatchery broodstock.		
TOTAL_BROODSTOCK_REMOVALS	Integer	This is the sum of adult and jack broodstock removals.		
OTHER_REMOVALS	Integer	Sexually maturing fish that have returned to the artificial / natural spawning grounds and were removed from the natural environment, by humans, prior to spawning for purposes other than collection of gametes. This includes in-river fisheries and surplus hatchery removals (fish that were initially removed for		

Field Name	Type	Description	Destination Table	Destination Field
		enhancement purposes but were not used for enhancement).		
TOTAL_RETURN_TO_RIVER	Integer	The complete accounting of sexually maturing fish that have returned to the freshwater environment. Total return to river = natural spawners + artificial spawners (e.g. hatchery broodstock) + other removals (harvest, ESSR).		
UNSPECIFIED_RETURNS	Integer	Sexually maturing fish that have returned to the freshwater environment. It is unknown whether this estimate refers to adults or adults and jacks, or whether it refers to the total return to river or a portion of. This is the field occupied by nuSEDS V1.0 estimates and some imported data. It is not a category available to estimates created after 2001.		
ENUMERATION_METHOD1	Text	The enumeration method used to observe fish. Values are: Bank Walk, Based on Angling Catch, Biologist/Working Group, Boat, Broodstock Removal, Dead Pitch, Electronic Counters, Electroshocking, Enumeration by Hatchery, Fence, Fixed Wing Aircraft, Float, Helicopter, Hydroacoustic Station, Other, Peak Live and Dead Count, Redd Counts, Snorkel, Spot Checks, Stream Walk, Strip Counts, Tag Recovery, Trap, Walk.		
ENUMERATION_METHOD2	Text	If more than one enumeration method was used.		
ENUMERATION_METHOD3	Text	If more than two enumeration methods was used.		
ENUMERATION_METHOD4	Text	If more than three enumeration methods was used.		
ENUMERATION_METHOD5	Text	If more than four enumeration methods was used.		
ENUMERATION_METHOD6	Text	If more than five enumeration methods was used.		
NATURAL_ADULT_FEMALES	Integer	Fraser River Specific field- All female salmon that have reached maturity, excluding jills (jills are female salmon that have matured at an early age).		
NATURAL_ADULT_MALES	Integer	Fraser River Specific field- All male salmon that have reached maturity, excluding jacks (jacks are male salmon that have matured at an early age).		
EFFECTIVE_FEMALES	Integer	Fraser River Specific field- Is the number of females estimated to have successfully spawned. This is calculated by multiplying the total female estimate (less removals eg. for fecundity estimation) by the weighted percent spawn.		
WEIGHTED_PCT_SPAWN	<NA>	Fraser River Specific field-Created to accommodate Fraser Sockeye data. This is based on evaluations of the success of spawn (fully		

Field Name	Type	Description	Destination Table	Destination Field
		spawned - 100%; partially spawned - 50%; unspawned - 0%) of individual female carcasses. A daily weighted percent spawn is applied to the total female recoveries for that day. The total effective females across the recovery period are summed and divided by the total female recoveries to calculate the overall weighted percent spawn.		
OTHER_ADULT_REMOVALS	Integer	created for South Coast (Somass System Sockeye).		
OTHER_JACK_REMOVALS	Integer	created for South Coast (Somass System Sockeye).		
TOT_ADULT_RET_RIVER	Integer	created for South Coast (Somass System Sockeye).		
TOT_JACK_RET_RIVER	Integer	created for South Coast (Somass System Sockeye).		
JUV_PRES_TYP	Integer	An indication of whether smolts or fry were present during the inspection. This field is from Historical Data (pre-2001) from Aeas 11 to 27, and Fraser chinook/coho, all from 1995 to 2001 .		
ACT_ID	Integer	This is the primary key for the SEN.	[DATA Escapement]	Id
POP_ID	Integer	Population ID.	[DATA Escapement] & [DATA NCCStreams]	SpeciesId & IndexId
SPC_ID	Integer	Species ID.		
GFE_ID	Integer	Stream ID (Geo_Feature ID).		
CREATED_DTT	Text	The date the SEN was created.		
UPDATED_DTT	Text	The date the SEN was updated.		

APPENDIX E. DERIVING THE MASTER NCC STREAM LIST

The master list of NCC streams is contained in the [DATA NCCStreams] table (Table 3-5) and is derived from the "Conservation Unit System Sites" file distributed with the NuSEDS data available at open.canada.ca[2]. This file contains a list of streams along with area and status designations (e.g., conservation unit and qualified species). This list was used as the basis for [DATA NCCStreams] based on the following steps:

1. Cross-reference SPECIES_QUALIFIED and POP_ID fields against the NuSEDS data to retrieved stream escapement data and determine the Area designations (i.e., the AREA field in NuSEDS);
2. Limit to streams designated as occurring in areas 1-10 (i.e., 1, 2E, 2W, 3A, 3B, 4A, 4B, 4C, 4D, 5, 6, 7, 8, 9, and 10);
3. Determine if the remaining streams have any recorded escapement records, if so mark the stream as "Active", only these streams were used in escapement calculations (e.g., Expansion Factor 2, see Section 5.3)
4. Drop record for lake type Sockeye for POP_ID = 42610, as the System does not distinguish between lake and river type Sockeye and both types occur for that POP_ID, which would lead to duplicate records in the System;
5. Apply the list of CU_INDEX changes (Table E-1) which revert conservation unit designations to those used previously within the System;
6. Apply the POP_ID corrections identified by Bruce Baxter on July 19, 2018 (Table E-2); and
7. Derive the internal fields outlined in Table 3-2.

After these steps the resulting data is imported as-is into [DATA NCCStreams] (Table 3-5). These steps are implemented by the `ImportNuSEDS()` and `UpdateNCCStreams()` routines (see Appendix C), which expect the Conservation Unit System Sites data to be in the format described in Table E-3.

[2] Full URL for the April 17, 2018 data release: https://open.canada.ca/data/en/dataset/c48669a3-045b-400d-b730-48aafe8c5ee6

Table E-1. List of streams with CU Index changes and associated CU designation revisions.

IndexId	SpeciesId	POP_ID	CU_INDEX		CU Designation	
			Current	Revised	Current	Revised
CM_51771	CM	51771	17	16	CM_16	CM_17
CO_50286	CO	50286	30	28	CO_30	CO_28
CO_50291	CO	50291	30	28	CO_30	CO_28
SX_3044	SX	3044	L-21-09	L-21-10	SX_L-21-09	SX_L-21-10
SX_3055	SX	3055	L-21-09	L-21-10	SX_L-21-09	SX_L-21-10
SX_3171	SX	3171	R16	L-19-71	SX_R16	SX_L-19-71
SX_45442	SX	45442	L-21-09	L-21-10	SX_L-21-09	SX_L-21-10
SX_45447	SX	45447	L-21-09	L-21-10	SX_L-21-09	SX_L-21-10
SX_45497	SX	45497	L-21-09	L-21-10	SX_L-21-09	SX_L-21-10
SX_49339	SX	49339	L-21-09	L-21-10	SX_L-21-09	SX_L-21-10
SX_49349	SX	49349	L-21-09	L-21-10	SX_L-21-09	SX_L-21-10
SX_49374	SX	49374	L-21-02	L-21-08	SX_L-21-02	SX_L-21-08
SX_51610	SX	51610	L-15-03	L-15-02	SX_L-15-03	SX_L-15-02
SX_51635	SX	51635	L-15-03	L-15-02	SX_L-15-03	SX_L-15-02

Table E-2. List of POP_ID and GFE_ID corrections applied to the Conservation Unit System Sites file before deriving the master list of NCC Streams.

SpecieId	POP_ID Corrections		GFE_ID Corrections	
	Current	Corrected	Current	Corrected
PKo	1	7600	4000001	23741
PKo	25521	26025521	3552	
PKe	25521	26025521	3552	
SX	53141	44948	2357	
SX	53142	42450	1496	
SX	53143	42610	1528	
CN	53151	7567	54169	

Table E-3. Expected format for the Conservation Unit System Sites data.

Field Name	Type	Description	Destination Table	Destination Field
ID	Integer	A unique numeric code identifying the site. One site can have numerous populations	[DATA NCCStreams]	SITE_ID
MAP_LABEL	Integer	N/A	[DATA NCCStreams]	
GFE_ID	Integer	Numeric code identifying the waterbody. From NUSEDS with some additions and modifications. Same as Stream_Id listed above	[DATA NCCStreams]	GFE_ID
SYSTEM_SITE	Text	The name of the waterbody. Originally from NUSEDS but not necessarily the same. Name priority was BC gazette > BC provincial alias > DFO alias1 > DFO alias2. Same as Waterbody Name above	[DATA NCCStreams]	
GFE_TYPE	Text	N/A	[DATA NCCStreams]	
POPMAP	Integer	The POP_ID onto which this POP_ID is mapped.	[DATA NCCStreams]	
SPECIES_QUALIFIED	Text	Conservation Unit acronym used to describe the species of salmon for which the escapement estimate is for, e.g.: CK - Chinook Salmon CM - Chum Salmon CO - Coho Salmon PKE - Even Year Pink Salmon PKO - Odd Year Pink Salmon SEL - Lake Type Sockeye Salmon SER - River or Ocean Type Sockeye Salmon	[DATA NCCStreams]	SPECIES_QUALIFIED & SpeciesId
YLAT	Number	Location of the mouth of the waterbody if flowing, or the centroid if not.	[DATA NCCStreams]	YLAT
XLONG	Number	Location of the mouth of the waterbody if flowing, or the centroid if not.	[DATA NCCStreams]	XLONG
FAZ_ACRO	Text	Acronym of the freshwater adaptive zone	[DATA NCCStreams]	FAZ_ACRO
MAZ_ACRO	Text	Acronym of the marine adaptive zone	[DATA NCCStreams]	MAZ_ACRO
JAZ_ACRO	Text	Acronym of the joint adaptive zone	[DATA NCCStreams]	JAZ_ACRO
CU_NAME	Text	The assigned name of the Conservation Unit. Note that this name does not identify the species.	[DATA NCCStreams]	CU_NAME
CU_ACRO	Text	The acronym assigned to the CU	[DATA NCCStreams]	CU_ACRO
CU_LATITUDE	Number	Centroid of the CU	[DATA NCCStreams]	CU_LATITUDE
CU_LONGITUDE	Number	Centroid of the CU	[DATA NCCStreams]	CU_LONGITUDE
NUMBER_OF_SITES	Integer	The number of POP_ID's in the CU. The value of this field and the latitude and longitude fields following come from a pivot table on the "spp_site_pivots" worksheets and are the total counts (#sites) and averages of lat and long for the sites in the CU. If you make changes to the sites in any CU refresh the pivot table to update the values of these fields.	[DATA NCCStreams]	NUMBER_OF_SITES
CU_TYPE	Text	There are currently six CU types, i.e.,Current, Bin, VREQ[Bin], VREQ[Current], VREQ[Extirpated] and Extirpated based upon Blair Holtby's	[DATA NCCStreams]	CU_TYPE

Field Name	Type	Description	Destination Table	Destination Field
		Rev 4.0 Conservation Unit data refresh.		
CU_INDEX	Text	A code assigned to the CU that when prefixed by the species code becomes the CU_index. Originally, these codes were assigned in a roughly S to N geographic order. With each revision to the CUs that order has been lost.	[DATA NCCStreams]	IndexId
FULL_CU_IN	Text	N/A	[DATA NCCStreams]	
SBJ_ID	Integer	NuSEDS code: 1=Sockeye, 2=Coho, 3=Pink, 4=Chum, 5=Chinook	[DATA NCCStreams]	SBJ_ID
POP_ID	Integer	NuSEDS POP_ID field. Unique number used to identify a specific population of salmon defined by species, spawning stream and run-timing	[DATA NCCStreams]	PopId
IS_INDICATOR	Text	Is "Y" if this POP_ID has been identified by Area experts as an "indicator" population.	[DATA NCCStreams]	IS_INDICATOR[1]
OL_GRP_NM	Text	A name assigned to the SACC Outlook Group	[DATA NCCStreams]	OL_GRP_NM
OL_GRP_N	Integer	A number assigned to the SACC Outlook Group.	[DATA NCCStreams]	OL_GRP_N
AREA	Text	Area Description	[DATA NCCStreams]	AREA
ISENH	Text	Is "Y" if either SEP_ENH or NUSEDS_Enh is "Y". If NUSEDS_Enh is "UNK" and SEP_ENH is "N" then is "UNK". Otherwise is "N"	[DATA NCCStreams]	ISENH
COMMENTS	Text	Any comments. This field was not added until late Sept 2012 and contains comments pertaining to changes required by 10Sep2012 snapshot only. And, as of Apr 2013, comments related to rev.4 sites changes/additions/etc.	[DATA NCCStreams]	COMMENTS
GFE_ID_IN_NUSEDS	Text	Is "Y" if the GFE_ID appears in the Geo_Features table of NUSEDS	[DATA NCCStreams]	GFE_ID_IN_NUSEDS
POP_ID_IN_NUSEDS	Text	Is "Y" if the POP_ID appears in NUSEDS. Note that there should be a reference somewhere to the NUSEDS snapshot against which the verification was done.	[DATA NCCStreams]	POP_ID_IN_NUSEDS
CMNT	Text	Comment Describing the version of the CU data	[DATA NCCStreams]	CMNT
EFFECTIVE_DT	Text	Date that the Conservation unit data became effective in NuSEDS	[DATA NCCStreams]	EFFECTIVE_DT
WATERSHED_CDE	Text	45 digit hierarchical provincial code unique to the waterbody and its watershed	[DATA NCCStreams]	
NEWWATERSHEDCDE	Text	45 digit hierarchical provincial code unique to the waterbody and its watershed	[DATA NCCStreams]	

[1] This field may be out of date relative the indicators used in [LOOKUP IndicatorStreams] (Table 5-3).

APPENDIX F. FORMAT FOR CATCH DATA EXTRACTED FROM THE FOS DATABASE

Catch data extracted from the FOS database is separated by gear type: gillnet, seine and troll (Table F-1).

Table F-1. Catch data extracted from the FOS database.

Data Item	Field Name(s)	Data Type	Description	Destination Table	Destination Field
Estimate Type	ESTIMATE_TYPE	Text	Catch estimate type		
License Area	LICENSE_AREA	Text	License area (e.g., Area A seine)		
Opening Category	OPNG_CAT	Text	Opening category		
Opening Description	OPNG_DESC	Text	Opening description		
Statistical Week	STAT_WEEK	Text	Canadian statistical week	[DATA FOSCatch]	StatWeek
Fishing Date	FISHING_DATE	Date	Date when the catch occurred		
Management Area	MGMT_AREA	Text	Statistical area number (e.g., 01, 02, 03)	[DATA FOSCatch]	StatArea
Area Name	AREA_NAME	Text	Sub-area name (e.g., Area 2E, 3-7A)	[DATA FOSCatch]	SubAreaId
Hours Open	HRS_OPEN	Number	Number of hours the fishery was open		
Vessels Operating	VESSELS_OP	Number	Number of boats		
Catch by Species	SOCKEYE_KEPT SOCKEYE_RELD COHO_KEPT COHO_RELD PINK_KEPT PINK_RELD CHUM_KEPT CHUM_RELD CHINOOK_KEPT CHINOOK_RELD STEELHEAD_KEPT STEELHEAD_RELD ATLANTIC_KEPT ATLANTIC_RELD DOGFISH_KEPT DOGFISH_RELD STURGEON_RELD	Number	Catch kept and released by species	[DATA FOSCatch]	SpeciesId, N
	COMMENTS	Text	Comments		

APPENDIX G. FORMAT FOR AGE COMPOSITION DATA EXTRACTED FROM DFO DATABASES

Age data was extracted from DFO databases in order to populate tables summarizing age compositions by conservation unit (Table G-1), statistical area (Table G-2), for the Babine Sockeye (Table G-3) and Kalum Late-run Chinook (Table G-4). The input data is expected to be in the format outlined in

Table G-1. Format of age composition data by CU extracted from the DFO databases.

Field Name(s)	Data Type	Description	Destination Table	Destination Field
Species	Text	Species	[DATA Age CU]	Species
Sp-CU_Name	Text	Species and conservation unit	[DATA Age CU]	Sp-CU_Name
CU	Text	Conservation unit name	[DATA Age CU]	CU
HasAgeData	Text	Conservation unit has age data	[DATA Age CU]	HasAgeData
Age 2	Number	Age composition data. Age 2 results are only for Pink salmon, while Age 3-7 results are computed based on associated catch, older and younger catch is not considered.	[DATA Age CU]	Age 2
Age 3				Age 3
Age 4				Age 4
Age 5				Age 5
Age 6				Age 6
Age 7				Age 7
CUstoUse	Number	Conservation units to use	[DATA Age CU]	CUstoUse
CU2	Number	Substitute age composition data from a different conservation unit or statistical area	[DATA Age CU]	CU2

Table G-2. Format of age composition data by statistical area extracted from DFO databases.

Field Name(s)	Data Type	Description	Destination Table	Destination Field
StatArea	Text	Statistical area	[DATA Age StatArea]	StatArea
Species	Text	Species	[DATA Age StatArea]	Species
Age 2	Number	Age composition data. Age 2 results are only for Pink salmon, while Age 3-7 results are computed based on associated catch, older and younger catch is not considered.	[DATA Age StatArea]	Age 2
Age 3				Age 3
Age 4				Age 4
Age 5				Age 5
Age 6				Age 6
Age 7				Age 7

Table G-3. Format of Babine age composition data by broodyear.

Field Name(s)	Data Type	Description	Destination Table	Destination Field
BroodYear	Number	Brood year	[DATA Age Babine Broodyear]	BroodYear
Age 3				Age 3
Age 4	Number	Age composition data	[DATA Age Babine Broodyear]	Age 4
Age 5				Age 5

Table G-4. Format of age composition data for Kalum Late Chinook salmon.

Field Name(s)	Data Type	Description	Destination Table	Destination Field
FISCAL_YEAR	Number	Year	[DATA Age Kalum Late]	FISCAL_YEAR
Yr2				Yr2
Yr3				Yr3
Yr4	Number	Age composition data	[DATA Age Kalum Late]	Yr4
Yr5				Yr5
Yr6				Yr6
Yr7				Yr7