**Survey\_summary\_data.r**

**Purpose**  This function is used to import, organize and run the analyses on the survey data. The first part of the script is used to load functions and to pre-process the data into a usable form. The second part of the script is used to run the analyses, such as MW-SH relationship, Survey biomass estimates, Condition Factor models, etc.

**Version Control**  This is the original version Survey\_summary\_figures.r, it along with Survey\_summary\_data.r replace the annual survey summary r scripts.

Required packages PBSmapping, RColorBrewer (the locally derived functions have their own package needs as well)

**Locally Derived Functions**

1. import.survey.data.r
2. get.offshore.survey.r
3. import.hyd.data.r
4. getdis.r
5. shwt.lme.r
6. condFac.r
7. surv.by.tow.r
8. simple.surv.r
9. restratwp.r
10. survey.dat.r
11. sprSurv.r

**Section 1**

This section is used to load in the locally derived functions, data from SQL and/or flat files. Several steps in here needed to make all the data line up and have all the necessary column names and such.

**Section 2**

This section is huge and is where all the analyses take place. This is basically a big for loop through all of the banks that you specified, things that are calculated for each bank in here are:

1. Strata is checked
2. MW-SH model
3. Condition Factor
4. Survey Biomass
5. Survey Abundance
6. Clapper Abundance
7. Clapper Percentages
8. German Lined and Unlined survey information
9. Meat Counts
10. Seebox survey abundance , biomass, etc

**Section 3**

At the end of the file there is a number of options for saving the data. It’s a bit of a smoggle, but the point of it is to ensure we don’t overwrite the overall survey results (Survey\_all\_results.Rdata) with a subset of the data, but if interested in a particular bank that the results are output to a datafile somewhere so they can be used elsewhere.

We also return a number of the useful objects from the function, including:

1. survey.obj This is a list of all the survey results for all the banks requested.

Depending on the nature of the survey for a bank this object has different levels of complexity for each bank. It contains both the time series of numerous variables and also contains the SHF data for the banks. Note that for German Bank this is simply the unlined survey results up to 2008 (see merged.survey.obj for the post 2008 time series and lined.survey.obj for the SHF’s since 2008.

1. SHF.summary This is the SHF summary for each bank from the survey.obj
2. SS.summary This is the survey time series for each bank from survey.obj
3. CF.current This is the condition factor for each bank, also includes the meat

counts for the current year. Useful for spatial plots showing CF and meat count variability

1. cf.data The condition factor as calculated by the GAM (or GLM on Middle

Bank) model for each tow on a bank.

1. clap.survey.obj The clapper survey results for each bank
2. lined.survey.obj Careful with this object, only the SHF data should be used from it.

This contains time series data for German bank for the “lined” surevy tows but it doesn’t account for the repeated sampling design so only the 2011 numbers are valid from the time series. What is useful from this is the SHF data it contains since 2008

1. merged.survey.obj This only contains the correct time series data for German bank

for the “lined” survey tows (so data starts in 2008). It accounts for the repeat tow design for the bank and for not having done repeat tows in 2011. Doesn’t have SHF data(see lined.survey.obj for SHF since 2008, and survey.obj for it before 2008)

1. seedbox.obj The results for any seedboxes that were closed in the year of

interest, will return a time series which includes years in which the seedbox was opened.

1. strata.mis.match If any of the strata are re-assigned this will provide a summary of

any tows in which strata from the database were re-assigned for this analysis

***Argument(s)***

1. direct The working directory to put figures are from which to grab data. Default = "Y:/Offshore

scallop/Assessment/",

1. yr.start Start year for getting the data. 1984 is default, we haven't generally used older data
2. yr End year. Default is the current system year: as.numeric(format(Sys.time(), "%Y"))
3. surveys Which banks+survey to pull the data for. Defaults to all banks except Banquereau which

isn't yet supported. Options are "BBnspring" ,"BBsspring" ,"Gerspring", "Midspring", "Sabspring", "GBspring" ,"GBbsummer", "GBasummer"

1. survey.year If specified this will overwrite the above "yr" and select the final year of the data,

Default = NULL. This is mostly useful if preprocessed = T (that is you already have data

up to current year and don't want to re-run the processing step but want to create the

survey data only up to a specific year).

1. preprocessed If you have already run the function and have the database data saved but want to

recreate the processed survey data. Default =F

1. un.ID Your SQL database username: Default = un.ID (which I suggest you specify in your R

profile so it is loaded by default)

1. pwd.ID Your SQL database password: Default = pwd.ID (if specified in your R profile it will be

loaded by default)

1. season For the spring survey we need to identify that we don't have all the results in yet. When

running the spring survey set to "spring". If just running GBa and GBb you can set this to "summer. When summer survey is complete you can also set this to the default of "both". Basically this is simply used to determine name of saved results.