

FastrCAT

Nissa Ferm

2019-01-24



Abstract The `FastrCAT` package makes it easy to use temperature and salinity data collected by a SeaBrid FastCAT. The package provides functions to read in .up files and bind them into a single dataframe, provide QA/QC feedback, plot depth by temperature/salinity for each cast, and maps of sample intensity, temperature, and salinity.

Introduction

Oceanographic data generated from SeaBird FastCAT CTD's have been collected since 1995 under EcoFOCI. The file type which we can view and collate the data from is a SeaBird proprietary file type called an .up file. The .up file is a post process file generated from the SeaSoft software and only includes the data from the upward cast of the device. Since 1997 there have been various versions of a Perl script which will append important station header information to the .up file. The Perl script acts as an interface between the SeaSoft software and the MasterCOD .db3 file to extract the header information based on the BON/CALVET/TUCK number. Each station from a cruise has its' own .up file. This format isn't in a traditional tabular format and is not easily accessed by users. It has been determined that it would be useful to build an R package for the FastCAT data that starts with a function to aggregate each cruises .up files into a tabular format. The function can then be used to generate tabular data and as a QAQC process for the .up files. Then the tabular data be easily used to create plots and calculate indices. This R package would be a final step in processing .up files while out at sea. The steps would be: enter COD form data into MasterCOD, run Perl script and then the R functions from the command line.

```
library(FastrCAT)
```

Functions in FastrCAT

```
make_dataframe_fc()
```

```
plot_ts_fc()
```

```
map_fc()
```

```
fill_missing_stations()
```

```
to_ecodaat()
```

```
plot_time_series()
```

make_dataframe_fc()

This function writes a single data frame in .csv format to file containing oceanographic data collected by the FastCat during a cruise. The format and column naming conventions are specific to the needs of EcoDAAT. This is the primary function of the FastrCAT package and must be run prior to all other functions. All other functions depend on the data frame generated. A cruise summary report is generated and written to file. The cruise

summary is in .html format. The summary contains basic information about the data and some summary statistics. Plots of the average temperature and salinity profiles and a station map. It also provides feedback on which .up files might be problematic due to header information or anomolous temperature and salinity readings.

Usage

```
make_dataframe_fc("path/ to directory/ of .up files")
```

Shows the first four rows and the first seven columns of an example dataframe.

```
#>   CRUISE STATION_NAME HAUL_NAME FOCI_GRID   DATE   TIME   LAT
#> 1 DY15-05           1         2    GR135 2015-05-15 23:23:00 55.4135
#> 2 DY15-05           1         2    GR135 2015-05-15 23:23:00 55.4135
#> 3 DY15-05           1         2    GR135 2015-05-15 23:23:00 55.4135
#> 4 DY15-05           1         2    GR135 2015-05-15 23:23:00 55.4135
```

All the column names:

```
#> [1] "CRUISE"           "STATION_NAME"      "HAUL_NAME"
#> [4] "FOCI_GRID"        "DATE"              "TIME"
#> [7] "LAT"              "LON"               "DEPTH_BOTTOM"
#> [10] "DEPTH"            "PRESSURE"          "TEMPERATURE1"
#> [13] "SALINITY1"        "CONDUCTIVITY1"     "SIGMA_T"
#> [16] "INSTRUMENT"       "DIRECTORY"         "CONDUCTIVITY2"
#> [19] "FLOUROMETER"     "OXYGEN1"           "OXYGEN2"
#> [22] "PAR"              "PH"                "POTENTIAL_DENSITY"
#> [25] "SALINITY2"        "TEMPERATURE2"      "TRANS"
#> [28] "COMMENTS_SEACAT_CTD" "GEAR_NAME"         "NET"
```

MasterCOD Acquired Information

All header information is hand entered into MasterCOD from a COD form for each consecutive FastCAT deployment. Header information acquired by the PERL script has not been consistent across all years. Since 1997 theses have been the various headers acquired from MasterCOD: latitude, longitude, date, time, cruise name, station name, haul name, Foci grid name, bottom depth, and instrument type. The function as built handles all types of header information listed above. Latitude and longitude are converted from degrees minutes seconds to decimal degrees. Date and time are converted into a date-time format set with a UTC time zone. One common error is that no header information is found. This error is most commonly due to the FastCat data being processed prior to station information being entered into MasterCOD. The function FastCat_compile.R within the package FastrCAT will generate a warning which is in the Cruise Report.

Example Header Warning in Cruise Report

```
"1" "This file G:/SeaCatData/Processed/1DY10/CAL001.up has no header info needs to be reprocessed."
```

This warning is telling the user that all header information is missing and that it needs to be reprocessed after the information has been entered into MasterCOD. Once this has been done the function make_dataframe_fc() can be run again. Human errors that are common are handled with exceptions within the function. When a user does not enter a field into MasterCOD the named header line is blank. The function will enter an NA.

****Example Header line in .up file:****

@ Cruise: DY12-05 ("DY12-05" in the dataframe)

@ Cruise:" " (NA in the dataframe)

Another error is when the Foci-grid designation or the BON number is entered into the station_name. The function will find this type of error and place the Foci_grid name into the correct column and leave the Station name blank. In other cases the Station name has been station.haul. The function will check to see if this is the case and separate the two and place station in station and haul in haul. When one of these three are missing it will require some corrections. The solution at sea would be to make sure that all these fields are entered into MasterCOD and then re-run the function. For older files this would require generating a table from EcoDAAT of Cruise, Haul, and FOCI_Grid to look-up and replace all missing values.

Issues that can't be resolved by the make_dataframe_fc() are when Cruise names or Foci_grid names are misspelled, which can be an issue when filtering or manipulating the resulting data. These issues would need to be QAQC'd after the fact and corrected. Sometimes, time, which is entered into MasterCOD as a four digit string is a nonsense time such as 0860. When this is the case, the function tries to convert it into a date-time object and cannot which results in an NA. Date can also be entered in incorrectly, if the format is correct the conversion to a date-time object will be fine, but the date might be incorrect. This would have to be QAQC'd after the data has been compiled by the function.

Seasoft Binned Data

In each .up file the numerical data is found in tab delineated columnar format after *END*. Each row in the .up file represents a single integer of pressure. The SeaBird FastCAT takes readings 16 times a second. For all values within an integer of pressure, one can select an average or an interpolated average. Over the years there have been as few as four columns of data up to nine. First the function looks to see if there is any data. If there is no data then a warning will be in the Cruise Report.

No data warning example from Cruise Report:

1" "This file G:/SeaCatData/Processed/DY17-05/BON055.up has no data associated with it."

When this occurs then header information is still gathered so there is a record of it but it will have blanks for all the numerical data columns. The function can handle the inconsistent column amounts. It looks for the name of each column and renames it in a standard format so that all data generated can be easily bound together. The data columns selected were pressure, temperature, conductivity, salinity, sigma_T, and flag. When there are more than these they will be dropped from the dataframe. Depth is added as a calculated value from pressure. The calculation is based on the AN69: Conversion of Pressure to Depth documented on the SeaBird website. This conversion requires latitude, so if latitude is missing depth will be an NA. Finally there are sometimes bad spikes in the data, a common cause being a bad termination. There is a filter in the function to remove temperatures less than -3 and greater than 20. As well as salinity values less than 0.05 and greater than 38.

Cruise Summary .html

Provides a short paragraph of number of tows, salinity, temperature, and location information. Summary statistics: minimum, median, mean, max, and number of NA's are also provided. This provides the user a paragraph ready for a cruise report and can be used as a quick data check. There is a plot of the average temperature salinity plot profile with 95% confidence intervals. A station map. And QA/QC information about missing header data, no data files, and files which contained outlier salinity and temperature data which was filtered out. If data seems incorrect then it is advised that values be corrected in MasterCOD and then to re-run the perl script and make_dataframe_fc(). Then a new summary will be generated.

Example of Cruise Summary

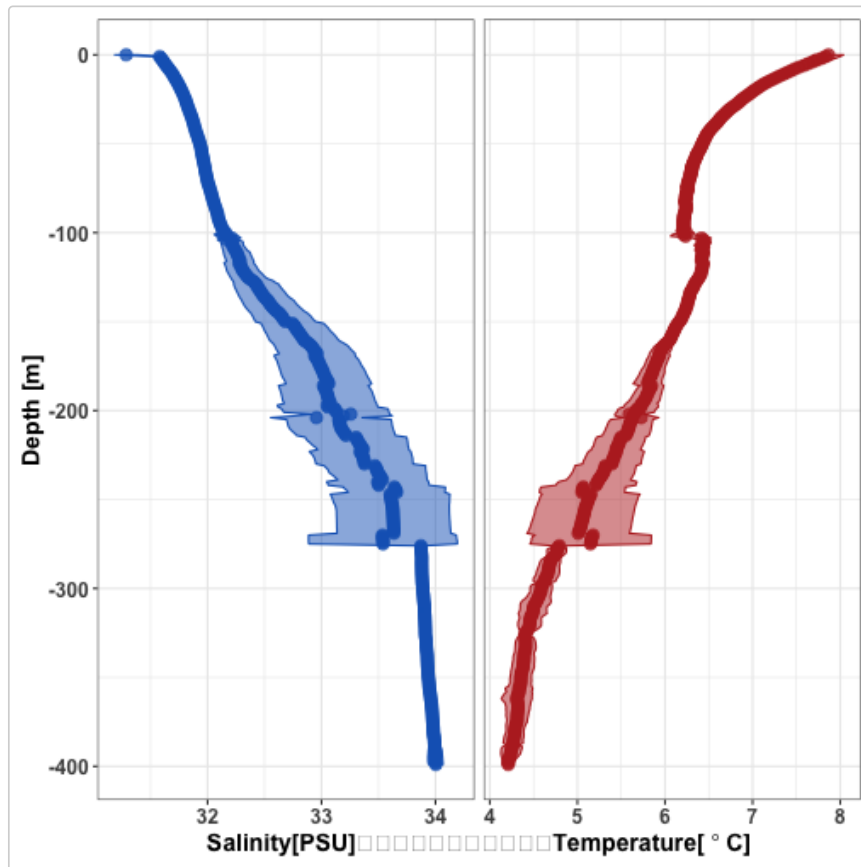
Cruise Report for DY15-05

Quick Cruise FastCAT Summary

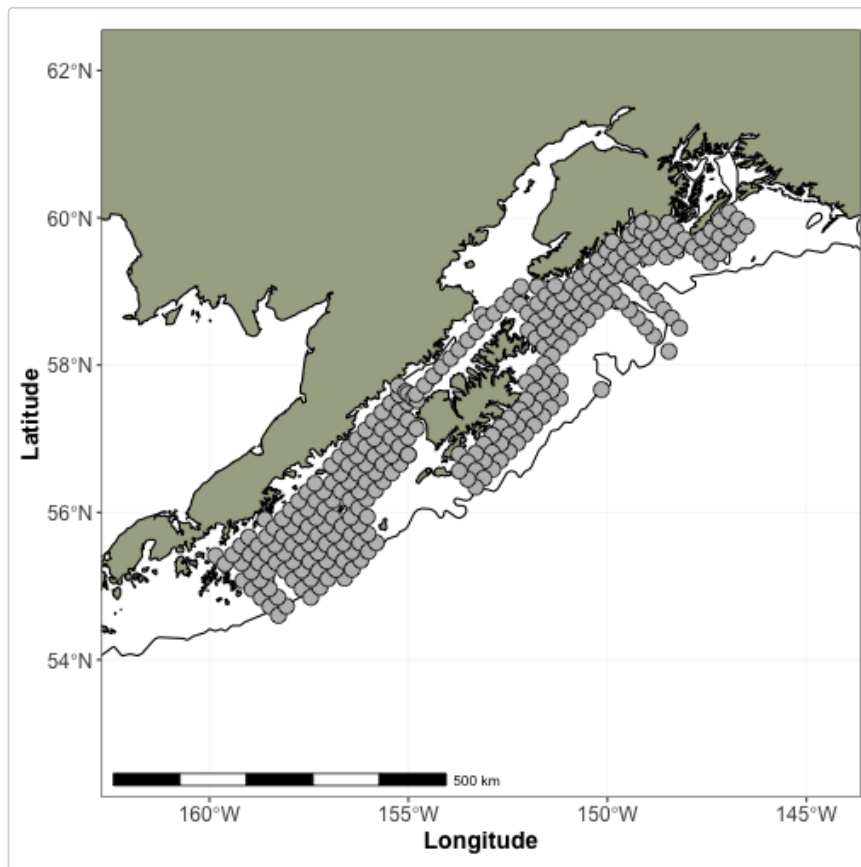
During the cruise DY15-05 there were 277 tows with depth, temperature, and salinity collected. Tows were taken from May 15 to June 4 of 2015. Temperature ranged from 4.2 to 12.8 with a mean of 6.6 Celcius. Salinity ranged from 25.6 to 34 with a mean of 31.97 PSU. The deepest tow was 400. The Southern most tow was 54.6 and the furthest North was 60.08 Latitude. The Western most tow was -159.85 and the Eastern most was -146.5 Longitude.

Average Salinity and Temperature Profiles

Plot shows average salinity and temperature (point) and 95% Confidence Intervals for each integer of depth.



Station Map



If any of this looks suspect, check for the values in the .csv file and then correct in MasterCOD. After this is done, re-run the Perl script and then re-run FastrCAT.

Summary Statistics

##	LAT	LON	DEPTH_BOTTOM	DEPTH
##	Min. :54.60	Min. :-159.8	Min. : 33.0	Min. : 0.00
##	Median :57.48	Median :-153.5	Median : 158.0	Median : 50.00
##	Mean :57.43	Mean :-153.5	Mean : 199.6	Mean : 58.04
##	Max. :60.08	Max. :-146.5	Max. :1625.0	Max. :400.00
##	PRESSURE	TEMPERATURE1	SALINITY1	CONDUCTIVITY1
##	Min. : 1.00	Min. : 4.176	Min. :25.62	Min. :27.50
##	Median : 51.00	Median : 6.559	Median :31.95	Median :32.16
##	Mean : 59.15	Mean : 6.594	Mean :31.97	Mean :32.20
##	Max. :405.00	Max. :12.760	Max. :34.04	Max. :35.86
##	SIGMA_T			
##	Min. :19.85			
##	Median :25.08			
##	Mean :25.08			
##	Max. :27.00			

List of Cruise Names If there is more than one Cruise name, please fix in MasterCOD.

```
## [1] "DY15-05"
```

List of Station.hauls

```
## [1] "1.2" "2.1" "3.1" "4.1" "5.1" "6.1" "7.1" "8.1"
## [9] "9.1" "10.1" "11.1" "12.1" "13.1" "14.1" "15.1" "16.1"
## [17] "17.1" "18.1" "19.1" "20.1" "21.1" "22.1" "23.1" "24.1"
## [25] "NA.1" "26.1" "27.2" "28.1" "29.2" "30.1" "31.2" "32.1"
## [33] "33.2" "34.1" "35.1" "36.1" "37.1" "38.1" "39.1" "40.1"
## [41] "41.1" "42.1" "43.1" "44.1" "45.1" "46.1" "47.1" "48.1"
## [49] "49.1" "50.1" "51.1" "52.1" "53.1" "54.1" "55.1" "56.1"
## [57] "57.1" "58.1" "59.1" "60.1" "61.1" "62.1" "63.1" "64.1"
## [65] "65.1" "66.1" "67.1" "68.1" "69.1" "70.1" "71.1" "72.1"
## [73] "73.1" "74.1" "75.1" "76.1" "77.1" "78.1" "79.1" "80.1"
## [81] "81.1" "82.1" "83.1" "84.1" "85.1" "86.1" "87.2" "88.1"
## [89] "89.2" "90.1" "91.3" "92.1" "93.3" "94.1" "95.3" "96.1"
## [97] "97.2" "98.1" "99.2" "100.1" "101.2" "102.1" "103.1" "104.2"
## [105] "105.2" "106.2" "107.2" "108.2" "109.2" "110.2" "111.1" "112.1"
## [113] "113.1" "114.1" "115.1" "115.2" "116.1" "117.1" "118.1" "119.1"
## [121] "120.1" "121.1" "122.1" "123.1" "124.1" "125.1" "126.2" "127.2"
## [129] "128.2" "129.2" "130.2" "131.2" "132.1" "133.1" "134.1" "135.1"
## [137] "136.1" "137.1" "138.1" "139.1" "140.1" "141.1" "142.1" "143.1"
## [145] "144.1" "145.1" "146.1" "147.1" "148.1" "149.1" "150.1" "151.1"
## [153] "152.1" "153.1" "154.1" "155.1" "156.1" "157.1" "158.2" "159.1"
## [161] "160.1" "161.1" "162.1" "163.1" "164.1" "165.1" "166.1" "167.1"
## [169] "168.1" "169.1" "170.1" "171.1" "172.1" "173.1" "174.1" "175.1"
## [177] "176.1" "177.1" "178.1" "179.1" "180.1" "181.1" "182.1" "183.1"
## [185] "184.1" "185.1" "186.1" "187.1" "188.1" "189.1" "190.1" "191.1"
## [193] "192.1" "193.1" "194.1" "195.1" "196.1" "197.1" "198.1" "199.1"
## [201] "200.1" "201.1" "202.1" "203.1" "204.2" "205.1" "206.1" "207.1"
## [209] "208.1" "209.1" "210.1" "211.1" "212.1" "213.1" "214.1" "215.1"
## [217] "216.1" "217.1" "218.1" "219.1" "220.1" "221.1" "222.2" "223.1"
## [225] "224.1" "225.1" "226.1" "227.1" "228.1" "229.2" "230.1" "231.1"
## [233] "232.1" "233.1" "234.1" "235.1" "236.1" "237.2" "238.1" "239.1"
## [241] "240.1" "241.1" "242.1" "243.1" "244.1" "245.1" "246.1" "247.1"
## [249] "248.1" "249.1" "250.1" "251.1" "252.1" "253.1" "254.1" "255.1"
## [257] "256.1" "257.1" "258.1" "259.1" "260.1" "261.1" "262.1" "263.1"
## [265] "264.1" "265.1" "266.1" "267.1" "268.1" "269.1" "270.1" "271.1"
## [273] "272.1" "273.1" "274.1" "275.1" "276.1"
```

List of FOCI grid names

```
## [1] "GR135" "GT135" "GX135" NA "HF141"
```

Anomalous Salinity and Temperature

These values for salinity and temperature are outside of known ranges. They have been filtered out of the data, because they are outside the know ranges for salinity and temperature. NA's refer to no outlier values found either below the minimum cutoff or above the maximum cutoff. If there is no data below, then all readings were within known ranges.

```
## [1] CRUISE STATION_NAME HAUL_NAME FOCI_GRID Min_value
## [6] Max_value ERROR_TYPE
## <0 rows> (or 0-length row.names)
```

No Header Files

These files have been identified as having no header information. It is most likely that you ran the Perl script prior to entering the station COD form into MasterCOD. Please re-run Perl script and then `make_dataframe_fc()` after you have checked that you entered the COD form. If there is not anything below then all header information was present.

```
## data frame with 0 columns and 0 rows
```

No Data Files

Below are files which were identified as having no data. Make sure that this is true. Since all casts are processed by the perl script and the `make_dataframe_fc()` regardless if they were Good, Questionable, or "Failure. If there is not a dataframe below then all files contained data.

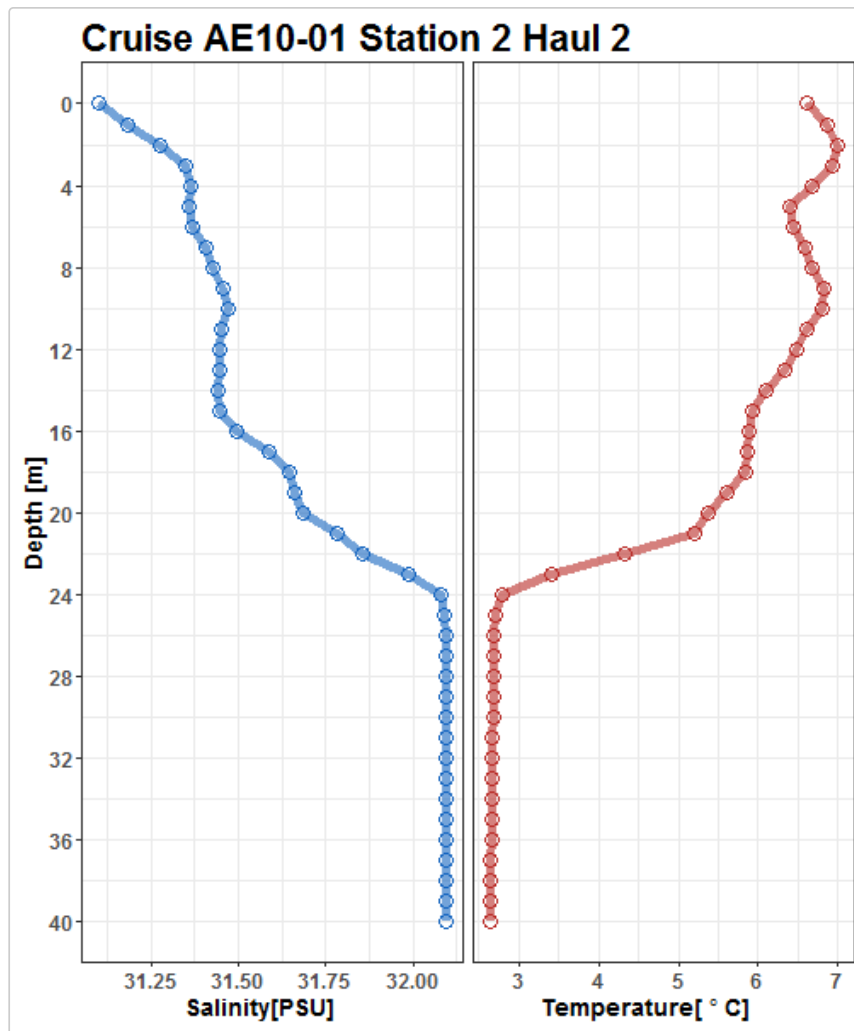
```
## data frame with 0 columns and 0 rows
```

plot_ts_fc()

Once `make_dataframe_fc()` has been run, then `plot_ts_fc()` can be used. This function creates a depth by salinity and temperature plot for each station.

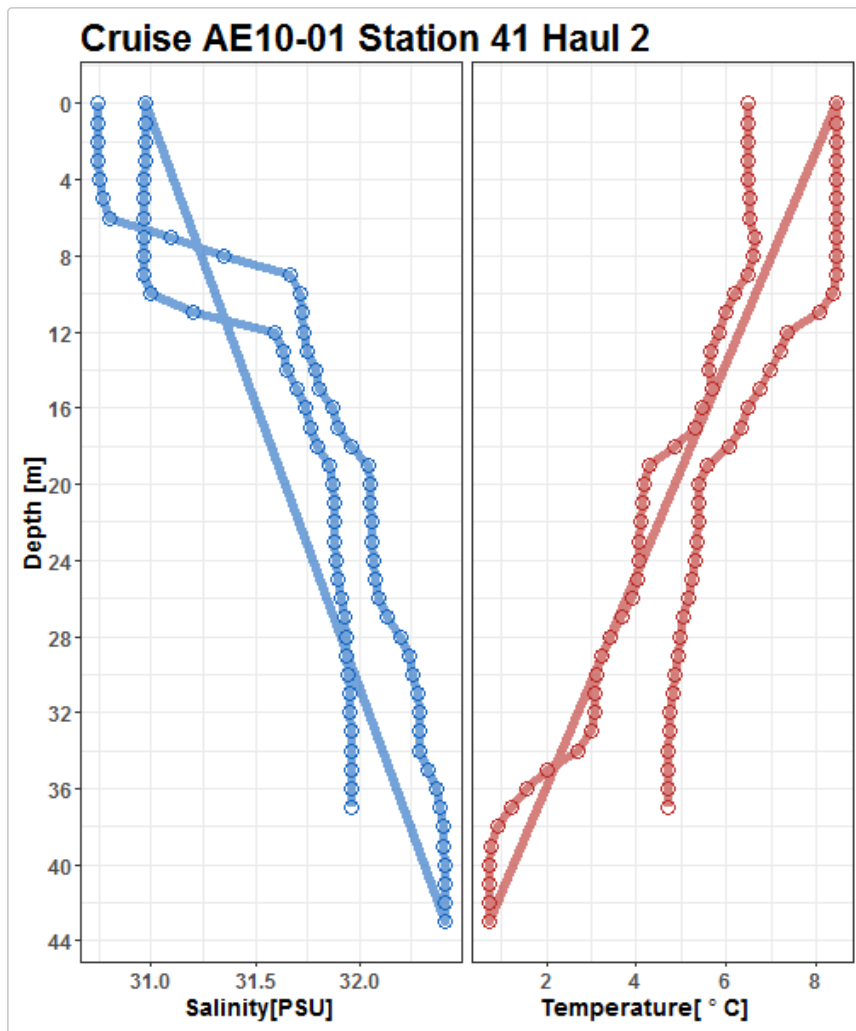
```
plot_ts_fc("path/ to directory/ of fastrcat dataframe")
```

These are all .png files which will be located in the plot folder within the current folder. It only needs to be run once to generate a plot for each station. Each dot is a data point. Check the profile for each station/haul.



Correct Output

For instance in the second plot below, it seems that two tows had the same Station and haul designation. If this occurs check the COD forms and MasterCOD and correct. This occurred because Station ID was station.haul and haul was not a value grabbed from MasterCOD. This allowed two consecutive casts to have the same station.haul information, where the second should have been Station 41, haul 3.



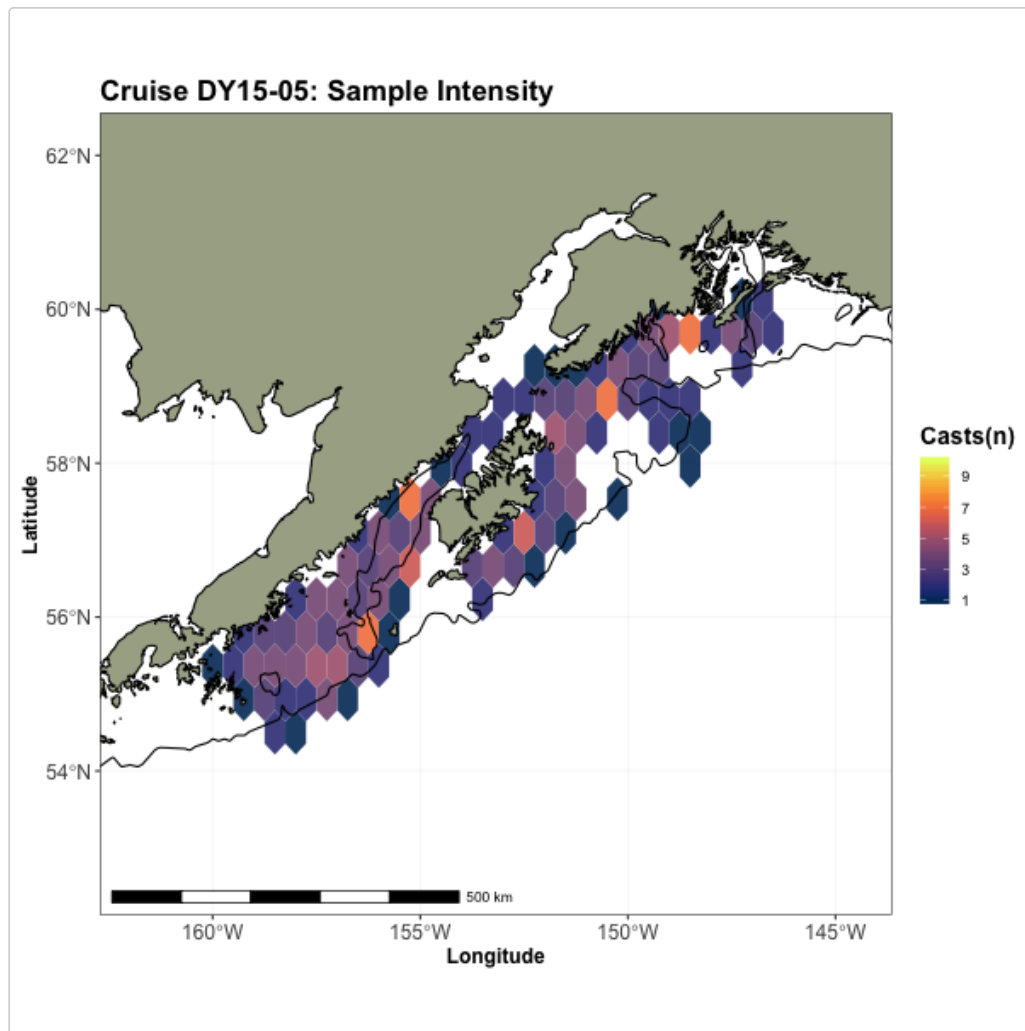
Incorrect Output

map_fc()

Once the `make_dataframe_fc()` has been run and the .csv file of the temperature, salinity, and conductivity is available you can use `map_fc()`. This function makes a station map by default. Other choices are sampling intensity, Salinity, and Temperature. The Salinity and Temperature maps are interpolated using inverse distance weighting and depth ranges can be selected as well.

```
map_fc("path/ to directory/ of fastrcat dataframe", map_type = "Sample Intensity")
```

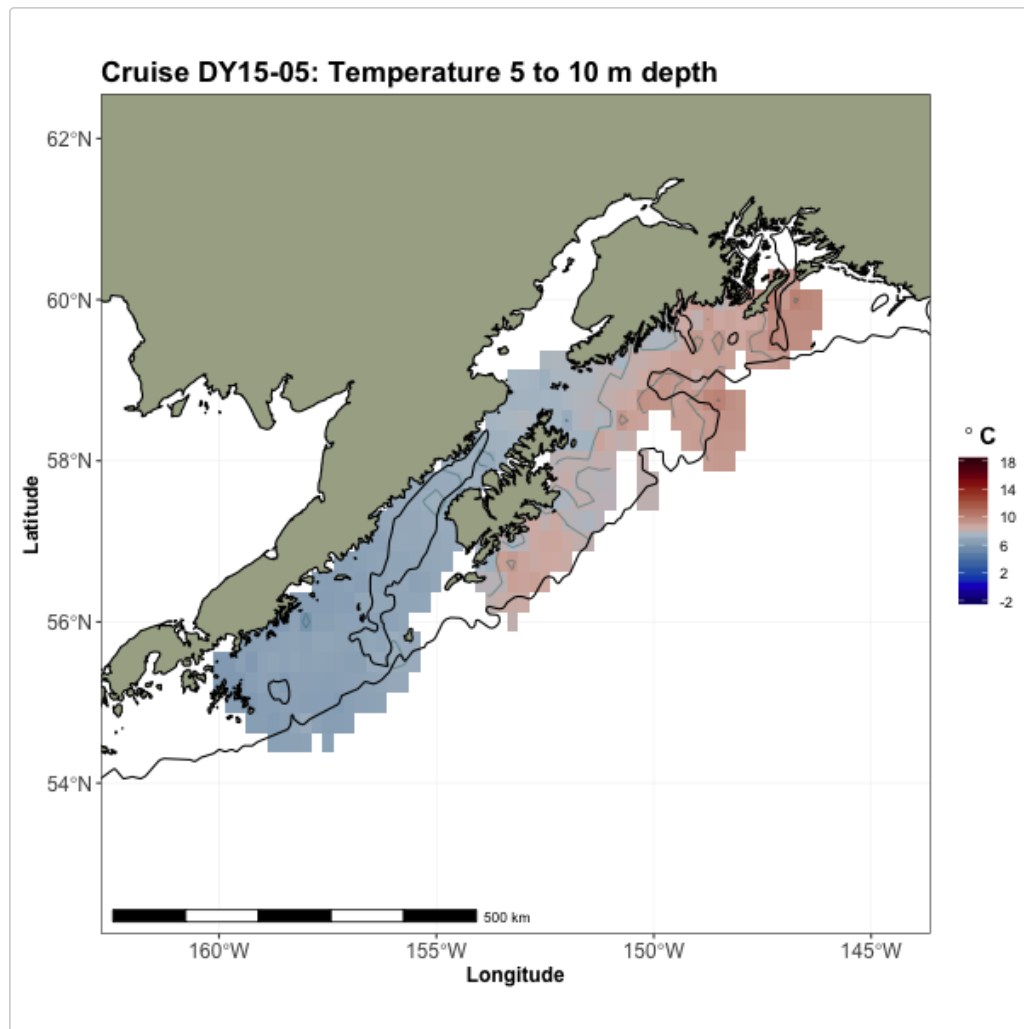
Here is an example of a sampling intensity map. Scales for legend will always be the same, for easy comparison among years.



Sample Intensity Map

Here is an example of a temperature map. This map shows the average temperature from 5 to 10 meters across the sampling grid.

```
map_fc("path/ to directory/ of fastrcat dataframe", map_type = "Temperature",  
depth_range = c(5,10))
```



Temperature Map

The following functions should be done once you have access to EcoDAAT with the corrected haul data and to past years data.

fill_missing_stations()

Like the name implies, this function will fill in the missing station name, foci grid name, and haul number in the dataframe which is generated by `make_dataframe_fc()`. This is a function which is used in conjunction with a dataframe in .csv format of haul records queried from EcoDAAT. This requires that the cruise haul information associated with the .up files has been uploaded into EcoDAAT. This function assists in data completeness after a cruise has been completed.

Usage

```
fill_missing_stations("path/ to directory/ of fastrcat dataframe",
                      "path/ to directory/ of haul records dataframe")
```

to_ecodaat()

This function preps data for easy loading into EcoDAAT. You provide it with a path to the folder with all the cruise seasons files prepared using the `make_dataframe_fc()` function. It will rename and bind together all files so only one file needs to be loaded into EcoDAAT per year. It is recommended that all QA/QC check have been done before and that the function `fill_missing_stations()` has been run to correct for missing haul information.

Usage

```
to_ecodaat("path/ to directory/of current years fastcat/ dataframes")
```

plot_time_series()

Use this function when you have historical and the current years data available. The function creates either temperature or salinity plots of core EcoFOCI stations in the Gulf of Alaska. Each plot displays the average temperature or salinity for each meter of depth of the core stations for each year for the months when peak sampling of these regions occurred. Line 8 and Semidi area are most commonly sampled in May and June, which is considered Spring. Summer sampling in the Gulf of Alaska has been less frequent and starts in the early 2000's. This summer sampling is in the Semidi core area, summer is considered August and September. Post 2010, these core stations were only sampled in odd numbered years. In the future more core areas will be added.

Usage

```
plot_time_series(hist_data = "path/ to directory/of historical .csv",
                 fastcat_data = "path to/ current years/fastcat .csv",
                 core_stations = "Line 8",
                 plot_type = "Temperature")
```

Using FastrCAT at Sea

Since people are already familiar with using the command line to run the Perl script out at sea it seemed logical to have people do the same for the R package functions. Requirements for running the package out at sea.

1. The most current version of R, R (32-bit) version 3.5.1 – “Feather Spray” must be installed on the at sea laptops.

2. All libraries and packages required to run functions must be installed. “list of dependencies”

3. The FastrCAT package

Example of Using FastrCAT in the command line:

1. **Change directory to where the R program files are.**

```
C:> cd C:/Program Files/R/R-3.5.1/bin
```

2. **Start an R session.**

```
C:Files-3.5.1> R
```

You will see a bunch info about R pop up. You can ignore this.

3. **Get the FastrCat package.**

```
> library("FastrCAT")
```

4. Run the function to make the data frame and cruise summary.

```
> make_dataframe_fc("path to current folder")
```

5. Run the function make all the temperature/salinity plots.

```
> plot_ts_fc("path to current folder")
```

6. Quit the R session.

```
> q()
```

7. It will ask if you want to save the workspace image. You don't need to.

```
> q()
```

```
Save workspace image? [y/n/c]: n
```

8. Now you are back into the command prompt. If you need to re-run the Perl script to make the .up files remember to change the working directory. Refer to the directions for setting up MasterCOD.

Comments on historical .up files

In order to re-run older files due to missing header information it will require finding the appropriate MasterCOD.db3 files and the folder with the .XMLCON, .hdr, and .hex files. The greatest amount of time will be spent retrieving the .db3 files for each cruise. An estimate for finding one file is 30 minutes. But once found it will only take a minute to reprocess. When evaluating the Processed folder it seems that 11 cruises will need to be reprocessed. There are most likely more cruises as the processed folder is incomplete. Some folders only have an excel file, this is the case for all Oshoro Maru cruises. This will require editing the excel file into the format requested for EcoDAAT.

<http://www.seabird.com/document/an69-conversion-pressure-depth>