SWEX Real-time Skew-T Plots

IMPORTANT: There are two important files:
/net/work/Projects/SWEX/real-time-plots/data_processing/
DO_NOT_DELETE_last_run_time
DO_NOT_DELETE_running
The date/time of the last_run_time file is used to determine which files in the ftp area are newer than when the process actually processed BUFR files the last time.

The running file is created when the process first starts, and then is moved to last_run_time file right before the process ends (only if BUFR files were processed during that run). The running file is used to determine if the process is already running, before it tries to process new files. Only one process should be processing files at one time.

If you want to rerun the process, the files in the FTP area must be touched, copied in again (without the -p option on the cp), or ftp'ed again. This is because the date/time of the files in the FTP area must be newer than the last_run_time file in order for the files to be processed.

There are additional documentation files here:

EOL-DMS/Skewt_plot_realtime/Real-time-Skew-T-Plotting:
Real-time-Skew-T-Plotting

EOL-DMS/Skewt_plot_realtime/Setting-Environment-for-Real-time-Skew-T-Plotting:
Setting-Environment-for-Real-time-Skew-T-Plotting

The main directory for this process is /net/work/Projects/SWEX/real-time-plots. The data_processing directory is where all the data files and logs are placed and where the two important files are kept. The software directory is where all the programs are located that run during this process.

Process will be run on tsunami as the joss user and in this order:

Crontab file on tsunami for the joss user:

0,10,20,30,40,50 * * * *
/net/work/Projects/SWEX/real-time-plots/software/Real-time-plots.pl

1. **GTS-BUFR files are received from source into an FTP area /net/ftp/pub/data/incoming/swex/radiosonde**

2. **Greg Stossmeister's process will rename the *.bufr files to the correct naming convention.** This process will only copy *.bfr files to be processed (see table below for the naming conventions). ????? Not sure if we need to do this for SWEX

Station Information:

Site	SiteID	Sonde Type	BUFR Message ID	TEMP Message ID	Naming Convention
Lake Cachuma Camp	ISS2	Vaisala	IUSB02		SWEX_ISS2_RS41_bufr309052_yyyymmdd_hhmmss.bfr
Santa Barbara Fire Department HQ	SBFD HQ	GRAW	IUSB01		SWEX_SBFDHQ_DFM09_bufr0309052_yyyymmdd_hhmmss.bfr
Santa Barbara Fire Station 18	SBFS 18	GRAW	IUSB04		SWEX_SBFS18_DFM09_bufr0309052_yyyymmdd_hhmmss.bfr
UCSB Sedgewick	ISS3	Vaisala	IUSB03		SWEX_ISS3_RS41_bufr309052_yyyymmdd_hhmmss.bfr

3. **Determine which files have not been processed and move to raw_gts_bufr_data directory under the data_processing directory.**
   - Copy *.bfr files only if DO_NOT_DELETE_running file does not exist.
   - Find *.bfr files that are newer than the last_run_time file.

4. **GTS-BUFR data conversion programs are run on files in data_processing/raw_gts_bufr_data.**

Here's what the plot generation software does:

- **Run conversion from Bufr to *.preproc then *.preproc to *.cls (ESC) files.** This calls the *preprocess_GTS_BUFR.pl* and *GTSBUFR_Radiosonde_Converter_RealTime.pl* software where sample calls to these programs would be

  preprocess_GTS_BUFR.pl ../raw_gts_bufr_data ../output_ascii_data ../output_preproc_data >& preproc_SWEX.log &

  And

  GTSBUFR_Radiosonde_Converter_RealTime.pl SWEX ../output_preproc_data ../output_esc 0 >& runSWEX_rt.log &

  (Beware that the preprocess_GTS_BUFR.pl s/w uses a BlackBox piece of s/w requiring specific library files. The latest of the bufr_dump s/w and lib is on eol-hurricane so you must execute this on that machine as of March 2025. You can learn the details of how the multi-step GTS-BUFR (dataset 100.030) data processing from the following links: DETAILED How To Process GTS BUFR Near Realtime Sounding Data for FIELD CATALOG PLOTS and How To: National Weather Service (NWS), GTS BUFR (100.030) Field Project Data Processing and Near Real Time Plot Generation and History .)

- **Run command to sort *.cls files.** This creates sorted *.cls files in directories in the data_processing area. Run the *ant sort_esc* command that can be found in the *build.xml* file as described in the linked documents above. To run the sort_esc command, ensure that ./src directory exists.

  (Running this sort command will generate the *record_sorter.log* file in the *../output_esc* directory. Once done, you will find the **.cls* and **.cls.unsort* files. **We plot the *.cls files.** Note that sorting the data may also cause any "0.00" values to be "-0.00" or vice versa depending on the machine you are running on.)

- **Warn if more than one sounding generated from GTS-BUFR file** (i.e., *_#.preproc or *_#.cls). Use the *_01_[SONDE|PIBAL].cls files for plotting.

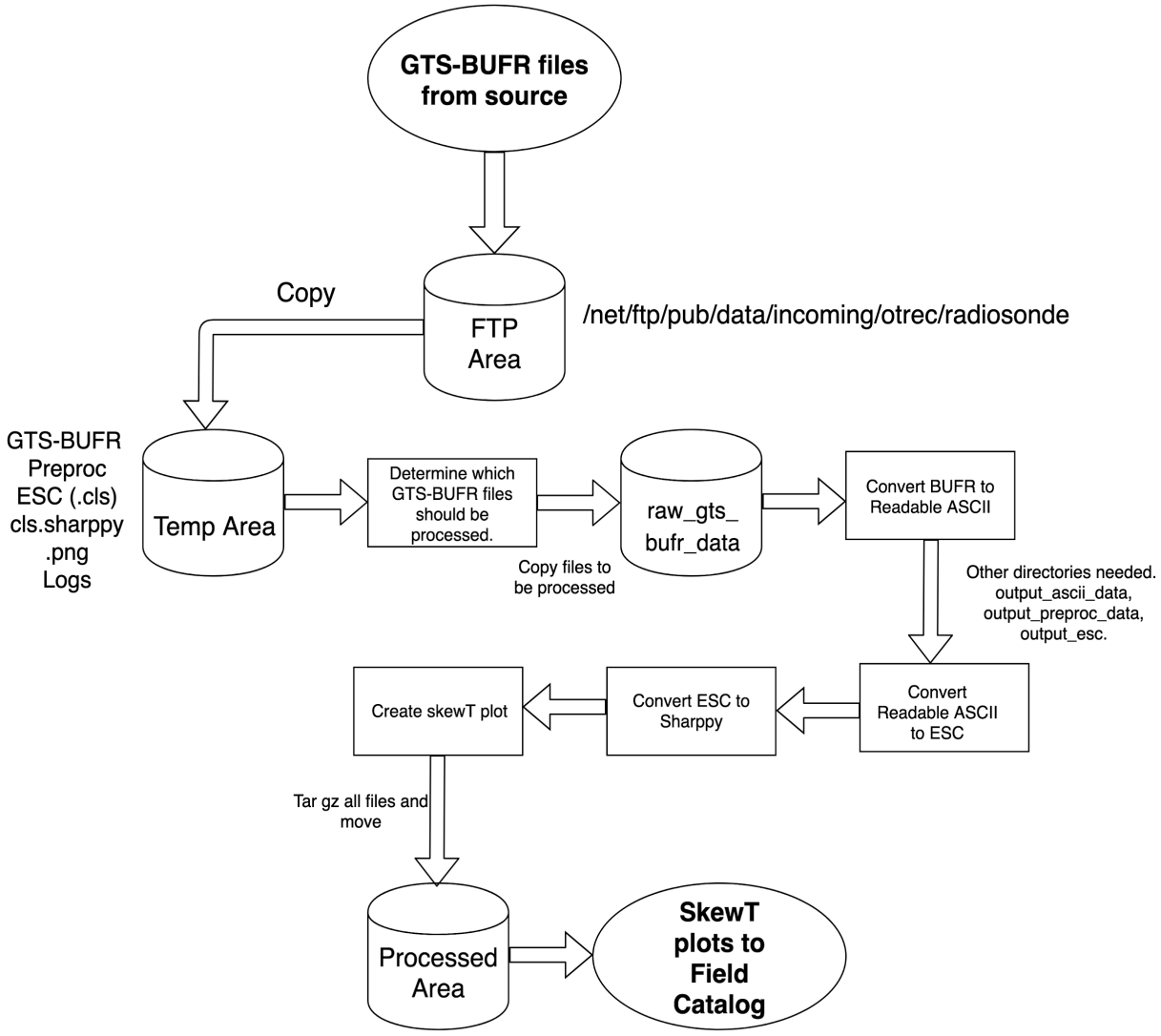
  (Note that more than one "preproc.cls" sounding file can be generated depending on what was included in the raw bufr data files. There can be many soundings in a single raw bufr input file and each will be given a successive two digit number in the output preproc.cls files.)

5. **Perl script (convertESCtosharppy.pl) is run to convert sorted *.cls files to .cls.sharppy format.**
    - Format that Sharppy reads.
    - Files created in Temp Area.

6. **Python script (plot_sounding.py) is run to create plots from .cls.sharppy files.**
    - Files are created in Temp Area.
    - .png plots are created.

7. **Tar/gzip all directories/files (including logs) and move to the processed area.**
    - Remove all files/directories from Temp Area except for DO_NOT_DELETE_last_run_time file.

8. Sounding Data Plots are copied to the Field Catalog area - /net/iftp2/pub/incoming/catalog/otrec
    - Naming convention, date/time used from header of cls file: upperair.Research_SkewT.YYYYMMDDhhmm.site.png



/net/iftp2/pub/incoming/catalog/otrec
upperair.Research_SkewT.YYYYMMDDHHmm.site.png

Issues/Comments

OTREC runs from August 5th to September 30th, 2019

Need a process to determine when a file has been completely written to the FTP area.

Temp Area is to be empty when there are no files to run.

Need to determine how often cron job will run to move GTS-BUFR files to the FTP area. The processing of the GTS-BUFR files needs to be complete before next move is done. Possible to check if Temp Area is empty before cron job runs again.

Duplicate named files - If considered an update to the previous file, then rerun by hand or make assume and have code overwrite/drop? Remove original files? Can plots be replaced in the field catalog? yes

Confirm testing of plot generation code if "junk" is generated from processing s/w. For instance, what if only a header is generated? What if no header and just junk text/chars? What else do we need to test through the whole process?

Eventually will need to tar, gzip files (by day?) because of file size.

Tested the following conditions:
Data not sorted in sounding - doesn't plot (generates errors).
If "Release Site" not included in header - Set to Unknown - Plots.
If "UTC Release Time" not included in header - Set to 000101/0000 (YYMMDD/HHmm) - Plots.

What If's

What if……

1. ***.bfr files are processed and have plots but *.bufr files are/do not?** Likely means the renaming script (run by Greg S.) did not work. Note that the data processing and plot generation s/w ONLY processes *.bfr input files.

2. **plot is not generated for a *.bfr file?** Possible reasons why…….

a. Could be junk (non BUFR) text in the *.bfr file or something else odd or the processing code can not interpret the data. Look at the log files in */net/work/Projects/OTREC/real-time-plots/data_processing/logs*. You may need to go into the /snapshot dir, match times in snapshot file name to find the one you want then gunzip and untar...then go to the /logs dir created to find the logs for that time. The overall process and each sub-process called generates a separate log file. First, look at the overall file (*otrec-rt-log-YYYYMMDDHHMM.log*) to see if everything ran (e.g., "find" cmd, cp bfr, run preprocess_GTS_BUFR.pl, run GTSBUFR_Radiosonde_Converter_RealTime.pl, "ant sort_esc", generate sharpy plots, cp skew-t's to catalog ingest area, tar/gzip files for future review, etc.). All log files show details about the exact calls. These calls could be executed at the command line by the joss user on tsunami.

b. Next, look at all the *output_** directories beneath */net/work/Projects/OTREC/real-time-plots/data_processing/.* If there are no *.bfr files in the raw* dir, then no output nor plots files would be created. If no input *.bfr files, then no output. If there are *.asc files in the output_ascii_data dir, then the first step of preprocess_GTS_BUFR.pl ran. If there are *.preproc and *.skiplines files in the output_preproc_data dir with a one for one , then the final step in preprocess_GTS_BUFR.pl ran. If there are *.cls files (but no *cls.unsort) files in the output_esc dir, then the unsorted cls files were created by GTSBUFR_Radiosonde_Converter_RealTime.pl but

3. **need to add another site or change the site information?** For OTREC, the site names, IDs, lat/lons, etc are hardcoded and you can find those and make updates in *GTSBUFR_Radiosonde_Converter_RealTime.pl* . Search for the word "HARDCODED" in that code or "LIMO" for the Limon OTREC site to find the others.