QCF to NetCDF Conversion Process

Updated 22 January 2025

This document, located here, lists the steps for running the QCF to Netcdf conversion process on a dataset. This spreadsheet, QCF Conversion to NetCDF, lists the datasets that were converted. The work for these conversions was done at */net/work/Projects/Convert_ASCII_to_netCDF* .

1) **Create a work area to run the conversions** under /net/work/Projects/Convert_ASCII_to_netCDF/[project]_[Frequency]Sfc/.

2) **Copy the *.qcf data files** for a specific dataset from the /net/archive/data location to your work area.

3) **Run the conversion program**, in the same directory where you copied the *.qcf files in your work area, as follows:

   /net/work/software/qcf2cdf/wrapper/convert-qcf-to-netcdf.pl *dataset-id* zithview

   dataset-id is the actual id of the dataset, i.e. 77.112. zithview is the database user. The program will ask for the zithview password which is **look-999**

   The program will find all *.qcf files in the current directory and run the correct conversion program on each data file.
   The output will look like the following for each *.qcf file in the current directory:

   ```
   Input file is hrly_311.qcf

   ncfile created is ./hrly_311_qcf.nc

   FileCopier done: total bytes written = 3569868, number of variables = 46
   ```

4) **Check one of the resulting *.nc files with the following:**
   ncdump -h file-name to see the header only

   ncdump file-name to see all the data in the file

   where file-name is one of the *.nc files created in step 3. Verify the following:
   1) cloud_level = 3 near the top of the output.
   2) The altitude variable is named **surface_altitude**. This is the 8th (or 7th) variable.

3) Verify the **flag values** for the **cloud_base_indicator_flag** variable near the end of the variables are listed like the following:
cloud_base_indicator_flag:flag_values = "U G M D B N X E C T I" ;
4) Verify the **title** under global attributes (near the bottom of the output when using the -h option) lists the correct dataset title.
5) Verify the **time_coverage_start** and **time_coverage_end** variables under global attributes (near the bottom of the output when using the -h option) list correct dates.

Hint: The best way to check the dates. From the command line:
bash (to use bash shell)
for f in *.nc; do echo $f; ncdump $f > $f.dump; done (All on one line)
exit (to return to your regular shell)
This will create a dump file for every .nc file, then run:
grep time_coverage *.nc.dump
Then all the dates can easily be checked. For dataset 1.33, the years are 1992, so this was the command used:
grep time_coverage *.nc.dump | grep -v 1992
to easily spot the bad dates.

Also, **spot check the data.** You can use **ncdump file-name > file-name.txt** where file-name.txt will contain a complete dump of the input file. Compare some values between the dump file and matching *.qcf file. **If you find anything that is not correct, please notify Janet, Linda C, Daniel.**

5) **Document your work in the DTS.**
As you work on a dataset, you need to document your work in the DTS. There may or may not be a project already in place in the DTS. Here's what to do in each case:

- **Look for the project in the DTS.**
  - **If the project and dataset are already in the DTS,** we can track our work using that existing DTS entry.
  - **If the project exists in the DTS but not the dataset,** then as normal we will add the dataset to the existing DTS project and track our work with this new entry.
  - **If neither project nor dataset exist in the DTS,** then we will add a dataset entry into the *Legacy Projects* project in the DTS. When you do this, make sure that you add the project name to the beginning of the dataset title like "WINTRE-MIX: xxxxx". This will help us keep the projects straight in the DTS Legacy Project.

6) **Load the *.nc data files into zinc.**
    a) Copy the *.nc data files into the **/net/archive/data** directory where the original *.qcf files are located. Find that location by looking in Zinc for the dataset. **Verify you are copying files that won't overwrite the original data files.** You do not need to create a new version of the dataset, these files will be part of the same dataset version.
    b) Update the **cksum** in the /net/archive/data/…. directory.
    c) Change the **dataset description** to include the following: ***"These data files are available as ASCII text files and netCDF data files."*** Note that sometimes the dataset may also have an E-BUFR version of the data in the archive that is orderable.
    d) Select the **OPeNDAP link** under Dataset. Under Actions on the OPeNDAP Setup page, click on Rebuild OPeNDAP auth files. **Note: There appears to be an issue with the OPeNDAP application. If the Rebuild OPeNDAP auth files button gives an error, Brooks may need to reload the OPeNDAP application.**
    e) Add the following **documentation files**:
    /net/archive/data/wmo-tables/WMO-0-20-003-present-weather.csv
    /net/archive/data/wmo-tables/WMO-0-20-011-cloud-amount.csv
    f) Add the **following link to the dataset**:
    https://library.wmo.int/records/item/35713-manual-on-codes-volume-i-1-international-codes
    This is xlink 6839.
    g) Verify that one of the *.qcf data files and one of the *.nc files correctly display in OPeNDAP.
    h) A document was created (by S. Loehrer) that discusses the differences in the units between the original ASCII QCF data and the newly created netCDF form of the data. This document (**Note_on_NetCDF_Data_File_Variable_Units.pdf**) must be loaded with each Surface netCDF dataset. **The following units changed between the ASCII QCF and the netCDF versions of the datasets:**
        1) Time and time_nominal - seconds since 1970-01-01.
        2) Surface_air_pressure, air_pressure_at_mean_sea_level, air_pressure_at_mean_sea_level_computed - converted from mbar to pascal.
        3) Air_temperature, dew_point_temperature - converted from celsius to kelvin.
        4) Precipitation_amount - units are noted as kg m-2 from mm although there was no conversion of the value necessary.

5) Cloud_base_altitude - converted from hundreds of feet to meters.

—-end of processing steps—

Additional Information

The following information is not needed for running the conversions.

1) There are 4 different conversion programs because there are different types of QCF surface data files.

QCFiosp.java - This conversion works for most of the QCF surface datasets.

QCFiospShortStn.java - This conversion is for datasets that have 10 characters for the station ID instead of the usual 15 characters. This applies to datasets 1.33 and 1.65.

QCFiospShortStnNoNominal.java - This conversion is for datasets that have 10 characters for the station ID instead of the usual 15 characters and also do not have the nominal time variable included. This applies to dataset 1.38.

QCFiospWind.java - This conversion is for datasets that have extra fields at the end of each record. This applies to datasets 19.014 and 40.010.

The wrapper script, convert-qcf-to-netcdf.pl, called the correct conversion for each dataset.

2) There were a few datasets that had bad dates in the data. The bad dates had 0011 for the year and others similar to that. For these data records, the following changes were made:

The year was changed to the same as the nominal year.
The minutes were changed to 55.
Any data quality flags that were G or D were changed to B.

This issue was noted in datasets 1.33 and 1.65.

3) All the software and documents for this can be found currently in svn: dmg/conversions/surface/QCF_surface2netcdf

—-end of document—