MET Pre-processing

George McCabe
(Originally created by Minna Win)

Developmental Testbed Center

Supported File Formats

- **Forecasts**
  - **GRIB1** – GRIdded Binary file
  - **GRIB2** – GRIB version 2 disabled by default (***--enable-grib2***)
  - **NetCDF** – Output from wrf_interp WRF-ARW utility, CF-Compliant versions 3 and 4, and internal MET NetCDF format
- **Gridded Analyses**
  - Same as Forecast file formats
  - GRIB Stage II/IV, MRMS, URMA, Model Analyses, GOES
  - **WWMCA** – World Wide Merged Cloud Analysis
  - **TRMM** – Tropical Rainfall Measuring Mission
  - **MODIS** – Moderate-Resolution Imaging Spectroradiometer
- **Point Observations**
  - **PREPBUFR** – binary data assimilation product (NDAS or GDAS)
  - **ASCII** – "MET specific" 11-column, little-r, SURFRAD, WWSIS, Aeronet
  - **MADIS** – METAR, Raob, Profiler, Maritime, Mesonet, or acarsProfiles
  - **LIDAR** - CALIPSO

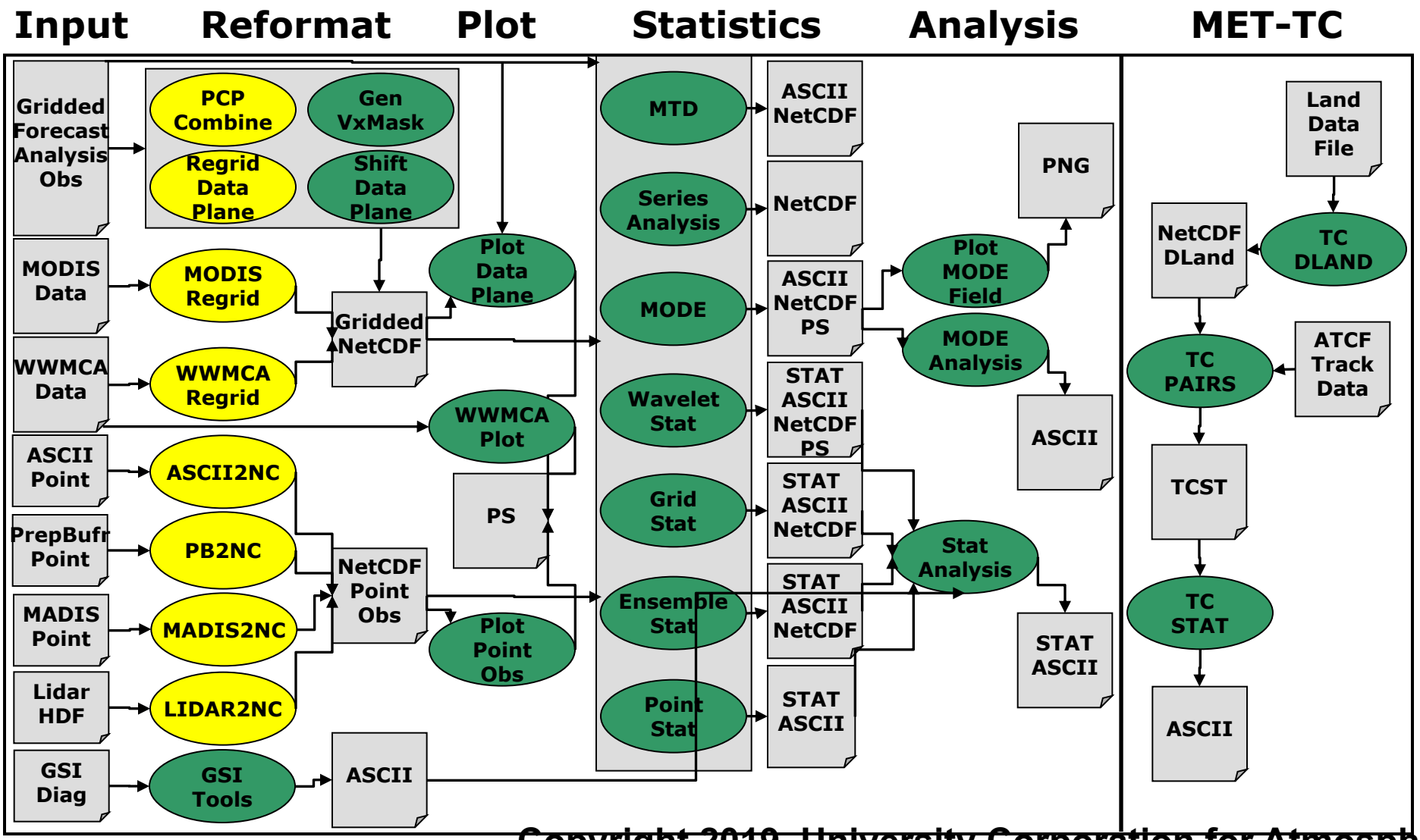
Data Inventory Tools

- **wgrib** – dumps GRIB1 headers and data.
  - http://www.cpc.ncep.noaa.gov/products/wesley/wgrib.html
- **wgrib2** – dumps GRIB2 headers and data.
  - http://www.cpc.ncep.noaa.gov/products/wesley/wgrib2/
- **ncdump -** dumps NetCDF headers and data.

  ncview – plots gridded NetCDF data.
  - http://www.unidata.ucar.edu/software/netcdf/

PRE-PROCESSING TOOLS

Pre-Processing / Reformatting



Data Reformatting Tools

- **PB2NC, ASCII2NC, MADIS2NC, LIDAR2NC**
  - Reformat point observations to the NetCDF format expected by Point-Stat and Ensemble-Stat.
- **MODIS-Regrid, WWMCA-Regrid**
  - Regrid HDF MODIS or binary WWMCA observations to the gridded NetCDF format expected by the MET statistics tools.
- **Regrid-Data-Plane**
  - Regrid one or more gridded data fields to user-specified grid.
- **PCP-Combine**
  - Add, subtract, or sum precipitation values across multiple gridded data files and write to the gridded NetCDF format expected by the MET statistics tools.

1. PB2NC Tool

- **Stands for "PREPBUFR to NetCDF"**
- **Functionality:**
  - Filters and reformats binary PREPBUFR and BUFR point observations into intermediate NetCDF format.
  - Configuration file specifies:
    - Observation types, variables, locations, elevations, quality marks, and times to retain or derive for use in Point-Stat or Ensemble-Stat.
- **Data formats:**
  - Reads PREPBUFR and BUFR using NCEP's BUFRLIB.
  - Writes point NetCDF as input to Point-Stat or Ensemble-Stat.

PB2NC

PREPBUFR

- **BUFR** is the World Meteorological Organization (WMO) standard binary code for the representation and exchange of observational data.
  - http://www.nco.ncep.noaa.gov/sib/decoders/BUFRLIB/
  - http://www.ecmwf.int/products/data/software/

- The **PREPBUFR** format is produced by NCEP for analyses and data assimilation. The system that produces this format:
  - Assembles observations dumped from a number of sources
  - Encodes
    - information about the observational error for each data type
    - background (first guess) interpolation for each data location
  - Performs both rudimentary multi-platform quality control and more complex platform-specific quality control
  - North American and Global datasets
- Only works with NCEP datasets with embedded tables.
- Support for external BUFR tables coming soon.

PB2NC: Usage

Usage: pb2nc
prepbufr_file
netcdf_file
config_file
[-pbfile prepbufr_file]
[-valid_beg time]
[-valid_end time]
[-nmsg n]
[-index]
[-dump path]
[-log file]
[-v level]
[-compress level]

prepbufr_file	**Input PrepBufr file name**
netcdf_file	**Output NetCDF file name**
config_file	**PB2NC configuration file**
-pbfile	**Additional input PrepBufr files**
-valid_beg -valid_end	**Beginning/Ending of valid time window** [YYYYMMDD_[HH[MMSS]]
-nmsg	**Number of PrepBufr messages to process**
-index	**Lists available BUFR variables**
-dump	**Dump entire contents of PrepBufr file to file in path**
-log	**Output file for log messages**
-v	**Level of logging**
-compress	**Compression level**

PB2NC: Run

- **pb2nc** ndas.t00z.prepbufr.tm12.20070401.nr \
  out/tutorial_pb.nc PB2NCConfig_tutorial -v 2

```
==> append : to filename to view the  data source
BUFR
230ADPUPA   UPPER-AIR (RAOB, PIBAL, RECCO, DROPS) REPORTS
231AIRCAR   MDCRS ACARS AIRCRAFT REPORTS                    232AIRCFT
AIREP/PIREP, AMDAR(ASDAR/ACARS), E-ADAS(AMDAR BUFR)
ACF233SATWND   SATELLITE-DERIVED WIND REPORTS
234PROFLR   WIND PROFILER REPORTS                           235VADWND
VAD (NEXRAD) WIND REPORTS                       236SATEMP   TOVS
SATELLITE DATA (SOUNDINGS, RETRIEVALS, RADIANCES) 237ADPSFC
SURFACE LAND (SYNOPTIC, METAR) REPORTS          238SFCSHP
SURFACE MARINE (SHIP, BUOY, C-MAN PLATFORM) REPORTS   239SFCBOG
MEAN SEA-LEVEL PRESSURE BOGUS REPORTS           240SPSSMI
SSM/I RETRIEVAL PRODUCTS (REPROCESSED WIND SPEED, TPW)
241SYNDAT   SYNTHETIC TROPICAL CYCLONE BOGUS REPORTS
242ERS1DA   ERS SCATTEROMETER DATA (REPROCESSED WIND SPEED)
243GOESND   GOES SATELLITE DATA (SOUNDINGS, RETRIEVALS,
RADIANCES) 244QKSWND   QUIKSCAT SCATTEROMETER DATA
(REPROCESSED WIND SPEED)   245MSONET   MESONET SURFACE
REPORTS (COOPERATIVE NETWORKS)      246GPSIPW   GLOBAL
POSITIONING SATELLITE-INTEGRATED PRECIP. WATER  247RASSDA   RADIO
ACOUSTIC SOUNDING SYSTEM (RASS) TEMP PROFILE
RPTSM063000BYTCNT                                   …
```

What obs are in a PREPBUFR file?

**>less \
ndas.t00z.prepbufr.tm12.
20070401.nr**

2. ASCII2NC Tool

- **Stands for "ASCII to NetCDF"**
- **Functionality:**
  - Reformat ASCII point observations into intermediate NetCDF format.
  - Multiple input ASCII formats supported (11-column, little-r, SURFRAD, WWSIS, and Aeronet).
  - Configuration file optional to define time summaries and message type mappings for little-r.
- **Data formats:**
  - Reads various input formats and writes point NetCDF as input to Point-Stat and Ensemble-Stat.
- ***Support for additional standard ASCII formats may be added as time and funding allow.***

ASCII2NC: Usage

Usage: ascii2nc
ascii_file
netcdf_file
[-format ascii_format]
[-config file]
[-mask_grid string]
[-mask_poly file]
[-mask_sid file|list]
[-log file]
[-v level]
[-compress level]

ascii_file	**Input ASCII file name**
netcdf_file	**Output NetCDF file name**
-format string	**met_point, little_r, surfrad, wwsis, aeronet**
-config file	**Optional configuration file name**
-mask_grid string	**Retain points within a named grid or gridded data file.**
-mask_poly file	**Retain points within a lat/lon polyline.**
-mask_sid file|list	**Retain a list of station ID's.**

ASCII2NC

MET-Point ASCII Format

Msg	STID	ValidTime	Lat	Lon	Elev	Var	Lvl	Hgt	QC	Ob	Ob assigns value to variable
ADPUPA	72365	20070331_120000	35.03	-106.62	1618.0	**7**	837.0	1618	NA	**1618**	***HGT**
ADPUPA	72365	20070331_120000	35.03	-106.62	1618.0	**11**	837.0	1618	NA	**273.05**	***TMP**
ADPUPA	72365	20070331_120000	35.03	-106.62	1618.0	**17**	837.0	1618	NA	**271.85**	***DPT**
ADPUPA	72365	20070331_120000	35.03	-106.62	1618.0	**52**	837.0	1618	NA	**92**	***RH**
ADPUPA	72365	20070331_120000	35.03	-106.62	1618.0	**53**	837.0	1618	9	**0.00417**	***MixRat**
ADPUPA	72365	20070331_120000	35.03	-106.62	1618.0	**7**	826.0	1724	2	**1724**	***HGT**
ADPUPA	72365	20070331_120000	35.03	-106.62	1618.0	**11**	826.0	1724	3	**274.55** …..	***TMP**

* *Use a value of "-9999" to indicate missing data*

Msg	**Message type**
STID	**Station ID**
ValidTime	**Valid time for observation**
Lat	**Latitude** [North]
Lon	**Longitude** [East]
Elev	**Elevation** [m] *(Note: currently not used by MET code so can be filled with -9999.)*
Var	**GRIB code or variable name** ***(i.e. AccPrecip or 61, MSLP or 2, Temp or 11, etc…)*** **http://www.cpc.ncep.noaa.gov/products/wesley/opn_gribtable.html**
Lvl	**Pressure** [mb] or **Accumulation Interval** [hr]
Hgt	**Height above Mean Sea Level** [m – MSL] *(Note: currently not used by MET code so can be filled with -9999.)*
QC flag	Quality control flag value
Ob	**Observed value**

ASCII2NC

MET-Point ASCII Format

Msg	STID	ValidTime	Lat	Lon	Elev	Var	Lvl	Hgt	QC	Ob
ADPUPA	72365	20070331_120000	35.03	-106.62	1618.0	**HGT**		837.0	1618	NA **1618**
ADPUPA	72365	20070331_120000	35.03	-106.62	1618.0	**TMP**		837.0	1618	NA **273.05**
ADPUPA	72365	20070331_120000	35.03	-106.62	1618.0	**DPT**		837.0	1618	NA **271.85**
ADPUPA	72365	20070331_120000	35.03	-106.62	1618.0	**RH**		837.0	1618	NA **92**
ADPUPA	72365	20070331_120000	35.03	-106.62	1618.0	**MIXR**		837.0	1618	9 **0.00417**
ADPUPA	72365	20070331_120000	35.03	-106.62	1618.0	**HGT**		826.0	1724	2 **1724**
ADPUPA	72365	20070331_120000	35.03	-106.62	1618.0	**TMP**		826.0	1724	3 **274.55**

* *Use a value of "-9999" to indicate missing data*

Msg	**Message type**
STID	**Station ID**
ValidTime	**Valid time for observation**
Lat	**Latitude** [North]
Lon	**Longitude** [East]
Elev	**Elevation** [m] *(Note: currently not used by MET code so can be filled with -9999.)*
Var	**GRIB code or variable name** ***(i.e. AccPrecip or 61, MSLP or 2, Temp or 11, etc…)*** **http://www.cpc.ncep.noaa.gov/products/wesley/opn_gribtable.html**
Lvl	**Pressure** [mb] or **Accumulation Interval** [hr]
Hgt	**Height above Mean Sea Level** [m – MSL] *(Note: currently not used by MET code so can be filled with -9999.)*
QC flag	Quality control flag value
Ob	**Observed value**

ASCII2NC: Run

- **ascii2nc** sample_obs.txt sample_ascii.nc -v 2

```
netcdf sample_ascii {
dimensions:
        mxstr = 15 ;
        hdr_arr_len = 3 ;
        obs_arr_len = 5 ;
        nhdr = 5 ;
        nobs = UNLIMITED ; // (2140 currently)
variables:
        char hdr_typ(nhdr, mxstr) ;
                hdr_typ:long_name = "message type" ;
        char hdr_sid(nhdr, mxstr) ;
                hdr_sid:long_name = "station identification" ;
        char hdr_vld(nhdr, mxstr) ;
                hdr_vld:long_name = "valid time" ;
                hdr_vld:units = "YYYYMMDD_HHMMSS UTC" ;
        float hdr_arr(nhdr, hdr_arr_len) ;
                hdr_arr:long_name = "array of observation station header values" ;
                hdr_arr:_fill_value = -9999.f ;
                hdr_arr:columns = "lat lon elv" ;
                … ;
        float obs_arr(nobs, obs_arr_len) ;
                obs_arr:long_name = "array of observation values" ;
                obs_arr:_fill_value = -9999.f ;
                obs_arr:columns = "hdr_id gc lvl hgt ob" ;
                obs_arr:hdr_id_long_name = "index of matching header data" ;
                … ;
```

Result of ncdump –h

Result of ncdump –v obs_arr

```
obs_arr =
0, 7, 837, 1618, 1618,
1, 11, 837, 1618, 273.05,
2, 17, 837, 1618, 271.85,
3, 52, 837, 1618, 92,
4, 53, 837, 1618, 0.00417,
5, 7, 826, 1724, 1724,
6, 11, 826, 1724, 274.55,
7, 17, 826, 1724, 272.15,
8, 52, 826, 1724, 84,
9, 53, 826, 1724, 0.00432,
10, 7, 815.3, 1829, 1829,
11, 11, 815.3, 1829, 276.45,
12, 17, 815.3, 1829, 265.75,
13, 52, 815.3, 1829, 45,
14, 53, 815.3, 1829, 0.0027,
15, 7, 815, 1832, 1832,
16, 11, 815, 1832, 276.55,
17, 17, 815, 1832, 265.55,
18, 52, 815, 1832, 44,
19, 53, 815, 1832, 0.00266,
20, 7, 784.7, 2134, 2134,
21, 11, 784.7, 2134, 274.05,
22, 17, 784.7, 2134, 264.15,
23, 52, 784.7, 2134, 47,
...
```

3. MADIS2NC Tool

- **Stands for "MADIS to NetCDF"**
- **Functionality:**
  - Reformat MADIS point observations into intermediate NetCDF format.
  - No configuration file.
- **Data formats:**
  - Reads MADIS METAR, RAOB, Profiler, Maritime, Mesonet, or acarsProfiles types.
  - Writes point NetCDF as input to Point-Stat or Ensemble-Stat.

MADIS2NC: Usage

Usage: madis2nc
madis_file
out_file
-type str
[-qc_dd list]
[-lvl_dim list]
[-rec_beg n]
[-rec_end n]
[-mask_grid string]
[-mask_poly file]
[-mask_sid file|list]
[-log file]
[-v level]
[-compress level]

madis_file	**Input MADIS NetCDF file name**
out_file	**Output NetCDF file name**
-type str	**metar, raob, profiler, maritime, mesonet, or acarsProfiles**
-qc_dd list	**QC flag values to be accepted (Z,C,S,V,X,Q,K,G,B)**
-lvl_dim list	**Vertical level dimensions to be processed**
-rec_beg n	**First MADIS record to process**
-rec_end n	**Last MADIS record to process**
-mask_grid string	**Retain points within a named grid or gridded data file.**
-mask_poly file	**Retain points within a lat/lon polyline.**
-mask_sid file|list	**Retain a list of station ID's.**

MADIS2NC: Run

- **madis2nc** profiler_20150409_1800.nc test.nc –type \
  profiler -v 2

```
DEBUG 1: Reading MADIS File: profiler_20120409_1800.nc
DEBUG 1: Writing MET File:   test.nc
DEBUG 2: Processing PROFILER recs     = 22
DEBUG 2: Rejected based on QC         = 0
DEBUG 2: Rejected based on fill       = 1674
DEBUG 2: Retained or derived = 1494
```

Result of ncdump –v obs_arr

```
obs_arr =
0 , 33 , -9999 , 1000 , -0.6316155 ,
0 , 34 , -9999 , 1000 , -0.9334552 ,
0 , 33 , -9999 , 1250 , -0.4383373 ,
0 , 34 , -9999 , 1250 , 1.078402 ,
0 , 33 , -9999 , 2250 , 1.004951 ,
0 , 34 , -9999 , 2250 , -0.9307967 ,
0 , 33 , -9999 , 2500 , 0.9661151 ,
0 , 34 , -9999 , 2500 , -1.082675 ,
0 , 33 , -9999 , 3750 , 6.587607 ,
0 , 34 , -9999 , 3750 , -8.664121 ,
1 , 33 , -9999 , 500 , 0.2172839 ,
1 , 34 , -9999 , 500 , -2.199575 ,
1 , 33 , -9999 , 750 , -0.242378 ,
1 , 34 , -9999 , 750 , -1.682394 ,
1 , 33 , -9999 , 1000 , 0.2787634 ,
1 , 34 , -9999 , 1000 , -1.51813 ,
1 , 33 , -9999 , 1250 , 2.726679 ,
1 , 34 , -9999 , 1250 , -1.324189 ,
1 , 33 , -9999 , 1500 , 4.239741 ,
1 , 34 , -9999 , 1500 , -1.897019 ,
1 , 33 , -9999 , 1750 , 3.581409 ,
1 , 34 , -9999 , 1750 , -5.975054 ,
```

8. LIDAR2NC Tool

- **Stands for "LIDAR to NetCDF"**
- **Depends on HDF4/HDFEOS libraries.**
- **Compilation disabled by default *(--enable-lidar2nc*)**
- **Functionality:**
  - Reformat LIDAR point observations into intermediate NetCDF format.
  - No configuration file.
- **Data formats:**
  - Reads CALIPSO Lidar data.
  - Writes point NetCDF as input to Point-Stat or Ensemble-Stat.
- ***Support for additional LIDAR formats may be added as time and funding allow.***

LIDAR2NC: Usage

Usage: lidar2nc

lidar_file

-out out_file

[-log file]

[-v level]

[-compress level]

lidar_file	**Input LIDAR HDF file name**
-out out_file	**Output NetCDF file name**

LIDAR2NC: Run

- **lidar2nc **

CAL_LID_L2_05kmCLay-Prov-V3-40.2016-12-01T01-24-58ZN.hdf \
  -out CAL_LID_L2_05kmCLay-Prov-V3-40.2016-12-01T01-24-58ZN.nc

```
DEBUG 1: Processing Lidar File:   data/lidar_data/CAL_LID_L2_05kmCLay-Prov-V3-40.2016-12-01T01-24-58ZN.hdf
DEBUG 1: Writing MET File:        tutorial/out/lidar2nc/CAL_LID_L2_05kmCLay-Prov-V3-40.2016-12-01T01-24-58ZN.nc
DEBUG 2: Processing Lidar points = 3728
```

```
obs_arr =
0, 500, _, 0, 1,
0, 501, 995.6906, 142.5747, 142.5747,
0, 502, 865.9296, 1160.472, 1160.472,
0, 503, 995.6906, 142.5747, 0,
0, 504, 995.6906, 142.5747, 100,
0, 601, 995.6906, 142.5747, 2,
0, 602, 995.6906, 142.5747, 0,
0, 603, 995.6906, 142.5747, 0,
0, 604, 995.6906, 142.5747, 3,
0, 600, 995.6906, 142.5747, 2,
0, 601, 995.6906, 142.5747, 2,
0, 505, 995.6906, 142.5747, 142.5747,
0, 506, 865.9296, 1160.472, 1160.472,
1, 500, _, 0, 1,
1, 501, 999.7334, 112.6365, 112.6365,
1, 502, 862.3634, 1190.41, 1190.41,
1, 503, 999.7334, 112.6365, 0,
1, 504, 999.7334, 112.6365, 100,
1, 601, 999.7334, 112.6365, 2,
1, 602, 999.7334, 112.6365, 0,
1, 603, 999.7334, 112.6365, 0,
```

Table 4.5: lidar2nc GRIB codes and their meaning, units, and abbreviations

GRIB Code	Meaning	Units	Abbreviation
500	Number of Cloud Layers	NA	NLayers
501	Cloud Layer Base AGL	m	Layer_Base
502	Cloud Layer Top AGL	m	Layer_Top
503	Cloud Opacity	%	Opacity
504	CAD Score	NA	CAD_Score
505	Minimum Cloud Base AGL	m	Min_Base
506	Maximum Cloud Top AGL	m	Max_Top
600	Feature Type	NA	Feature_Type
601	Ice/Water Phase	NA	Ice_Water_Phase
602	Feature Sub-Type	NA	Feature_Sub_Type
603	Cloud/Aerosol/PSC Type QA	NA	Cloud_Aerosol_PSC_Type_QA
604	Horizontal Averaging	NA	Horizontal_Averaging

LIDAR2NC

Questions?