File Formats and Pre-Processing

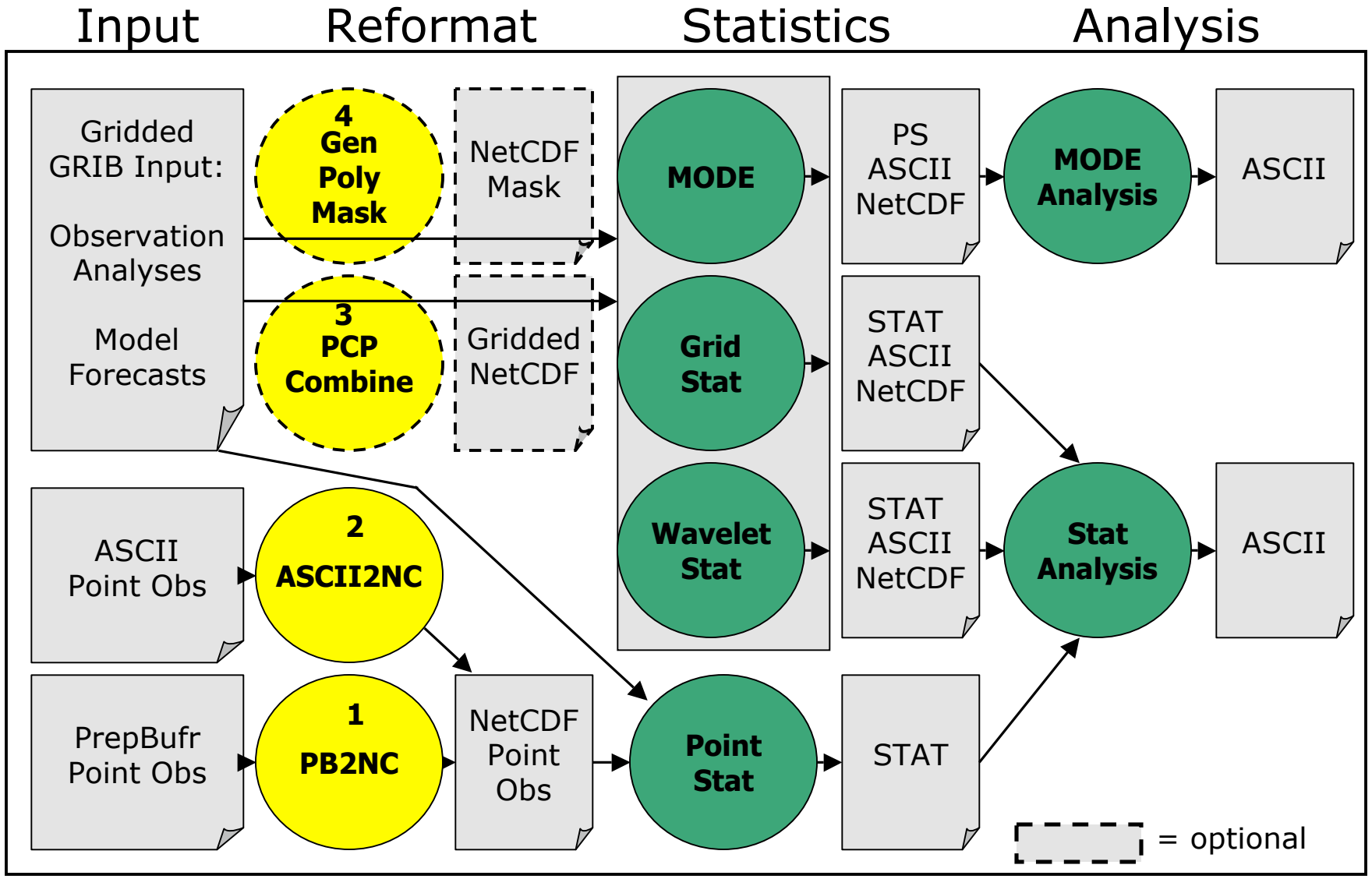
- File Formats
- Pre-processing Tools
- Useful Links

Presenter: Tara Jensen

Supported File Formats

- **Forecasts**
  - **GRIB** – GRIdded Binary file format (*version 1)
  - **NetCDF** – "MET specific" network Common Data Format
- **Observations**
  - **PREPBUFR** – binary dataset prepared by NCEP from varied data sources.
  - **ASCII** – "Met_Point" format (10-cols x n-rows)
  - gridded **GRIB** – i.e. NEXRAD Level II or IV
  - **NetCDF** – "MET specific"

Pre-Processing / Reformatting



Data Reformating Tools

- **PB2NC and ASCII2NC**
  - Arrange observational data into the NetCDF point format expected by Point-Stat.
- **PCP_Combine (optional)**
  - Sum precipitation values across two or more time periods.
  - Subtract precipitation values to create values for finer subperiods.
  - Produces gridded NetCDF file that can be used as input grid for any Statistics tool.
- **Gen_Poly_Mask (optional)**
  - Used when more complex masking is needed.
  - Produces a NetCDF file of pre-defined mask.
  - May be used for masking in any Statistics tools.

1. PB2NC Tool

- **Stands for "PREPBUFR to NetCDF"**
- **Functionality:**
  - Filters and reformats PREPBUFR 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.
- **Data formats:**
  - Reads PREPBUFR using NCEP's BUFRLIB.
  - Writes point NetCDF as input to Point-Stat.

Note: v2.0 no longer requires CWORDSH to pre-process PREPBUFR files.

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) interpolated to each data location
  - Performs both rudimentary multi-platform quality control and more complex platform-specific quality control
- **MET currently only supports PREPBUFR.** Let us know if you need to read in other BUFR format.

PREPBUFR - Known Issues

- CWORDSH blocking
  - In v1.0 and v1.1 it was needed to structure to data properly for reading by the MET fortran code.
  - Need eliminated in v2.0.
- Compilation 64-bit OS
  - Will work if you compile using 32-bit flags.
  - Investigation of this problem in ongoing.

PB2NC: Usage

Usage: pb2nc

prepbufr_file

netcdf_file

config_file

[-pbfile prepbufr_file]

[-valid_beg time]

[-valid_end time]

[-nmsg n]

[-dump path]

[-v level]

prepbufr_file	**Input obs file in PrepBufr format**
netcdf_file	**Output name for NetCDF file**
config_file	**PB2NC configuration file**
-pbfile	**Additional input PrepBufr files**
-valid_beg	**Beginning of valid time window** [YYYYMMDD_[HH[MMSS]]
-valid_end	**End of valid time window** [YYYYMMDD_[HH[MMSS]]
-nmsg	**Number of PrepBufr messages to process**
-dump	**Dump entire contents o PrepBufr file to file in path**
-v	**Level of logging**

PB2NC: Run

- **METv2.0/bin/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
...
```

Result of running linux "less" or "more" command:

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

2. ASCII2NC Tool

- **Stands for "ASCII to NetCDF"**
- **Functionality:**
  - Reformat ASCII point observations into intermediate NetCDF format.
  - One input ASCII format supported (10 columns):
  - No configuration file.
- **Data formats:**
  - Reads Met_Point ASCII.
  - Writes point NetCDF as input to Point-Stat.
  - ***Future: support additional standard ASCII formats.***

ASCII2NC: Usage

Usage: ascii2nc

ascii_file

netcdf_file

[-format ascii_format]

[-v level]

ascii_file	**Input obs file in ASCII format**
netcdf_file	**Output name for NetCDF file**
-format	**Override default MET_Point format** (*Future Option*)
-v	**Level of logging**

MET_Point Format

Msg	STID	ValidTime	Lat	Lon	Elev	GC	Lvl	Hgt	Ob
ADPUPA	72365	20070331_120000	35.03	-106.62	1618.0	7	837.0	1618	1618
ADPUPA	72365	20070331_120000	35.03	-106.62	1618.0	11	837.0	1618	273.05
ADPUPA	72365	20070331_120000	35.03	-106.62	1618.0	17	837.0	1618	271.85
ADPUPA	72365	20070331_120000	35.03	-106.62	1618.0	52	837.0	1618	92
ADPUPA	72365	20070331_120000	35.03	-106.62	1618.0	53	837.0	1618	0.00417
ADPUPA	72365	20070331_120000	35.03	-106.62	1618.0	7	826.0	1724	1724
ADPUPA	72365	20070331_120000	35.03	-106.62	1618.0	11	826.0	1724	274.55

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

Msg	**Message type**
STID	**WMO 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.)*
GC	**GRIB code for variable** ***(i.e. AccPrecip = 61; MSLP = 2; Temp = 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.)*
Ob	**Observed value**

ASCII2NC: Run

- **METv2.0/bin/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,
 0, 11, 837, 1618, 273.05,
 0, 17, 837, 1618, 271.85,
 0, 52, 837, 1618, 92,
 0, 53, 837, 1618, 0.00417,
 0, 7, 826, 1724, 1724,
 0, 11, 826, 1724, 274.55,
 0, 17, 826, 1724, 272.15,
 0, 52, 826, 1724, 84,
 0, 53, 826, 1724, 0.00432,
 0, 7, 815.3, 1829, 1829,
 0, 11, 815.3, 1829, 276.45,
 0, 17, 815.3, 1829, 265.75,
 0, 52, 815.3, 1829, 45,
 0, 53, 815.3, 1829, 0.0027,
 0, 7, 815, 1832, 1832,
 0, 11, 815, 1832, 276.55,
 0, 17, 815, 1832, 265.55,
 0, 52, 815, 1832, 44,
 0, 53, 815, 1832, 0.00266,
 0, 7, 784.7, 2134, 2134,
 0, 11, 784.7, 2134, 274.05,
 0, 17, 784.7, 2134, 264.15,
 0, 52, 784.7, 2134, 47,
```

3. PCP-Combine Tool

- **Stands for "Precip-Combine"**
- **Functionality:**
  - Mathematically combines precipitation fields across multiple files.
    - Add precipitation over 2 files
      - *2 NMM output files to go from 3-hr to 6-hr accumulation.*
    - Sum precipitation over more than 2 files
      - *12 WSR-88D Level II data to go from 5 min accumulation to 1-hr accumulation.*
    - Subtract precipitation in 2 files
      - *2 ARW output files to go from 12 hr accumulations to 6 hour accumulation*
  - Specify field name on the command line.
  - No configuration file.
- **Data formats:**
  - Reads GRIB format.
  - Writes gridded NetCDF as input to stats tools.

PCP-Combine: Usage

Usage: pcp_combine

[-sum sum_args]

or [-add add_args]

or [-subtract sub_args]

[-gc code]

[-v level]

-sum	**Accumulates data over multiple files.** ***Sum_args***: (init_time, in_accum, valid_time, out_accum, out_file, -pcpdir path, -pcprx reg_exp)
-add	**Accumulates data over two files.** ***Add_args:*** (in_file1, Accum1, in_file2, Accum2, out_file).
-subtract	**Subtracts data over two files.** ***Sub_args:*** (in_file1, Accum1, in_file2, Accum2, out_file).
-gc	**GRIB code for variable** *(i.e. ACPC = 61; NCPCP =62; ACPCP =63, etc…).*
-v	**Level of logging**

PCP-Combine: Run

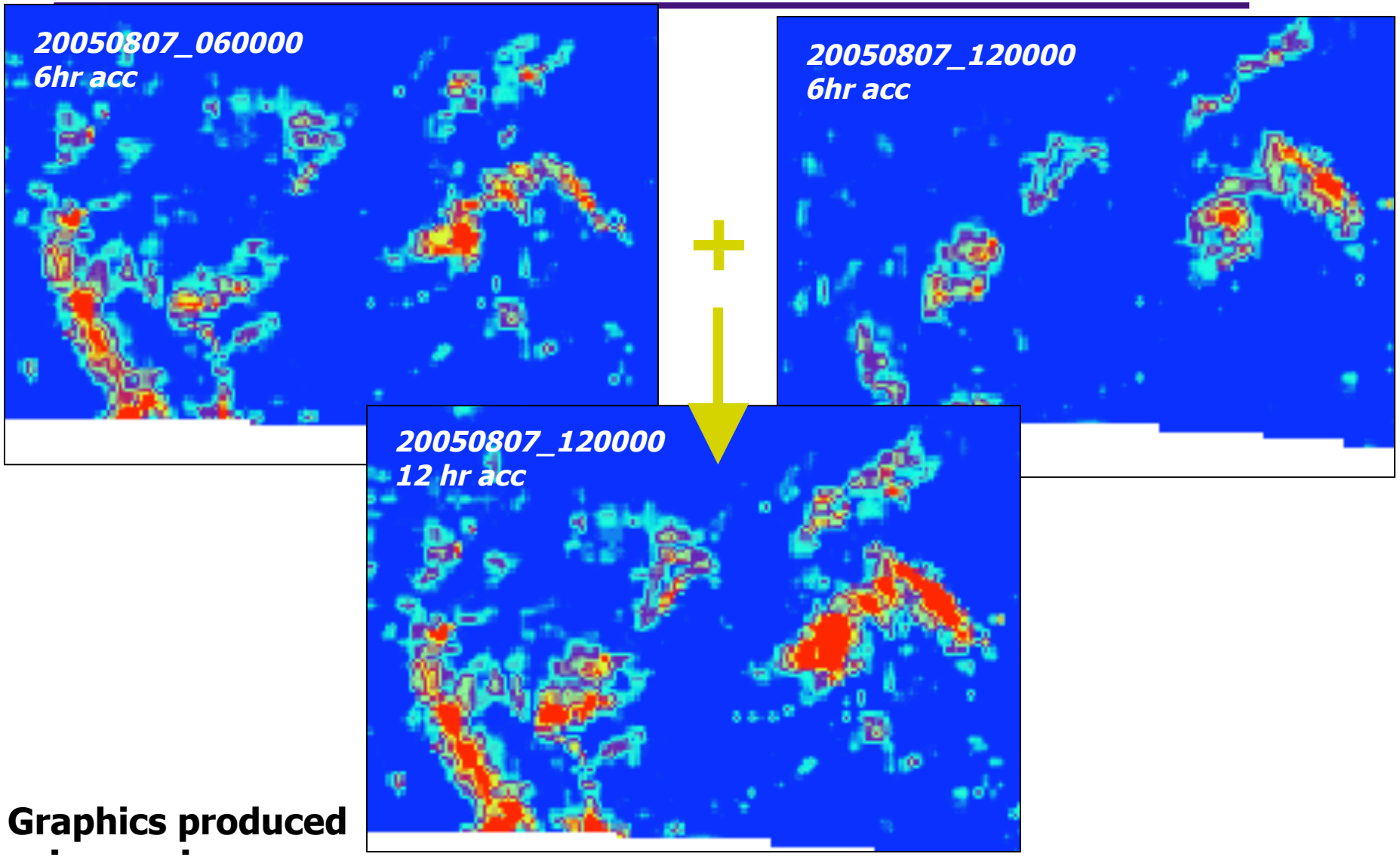
- **Two example command lines**
  - Summing two 6-hourly accumulation forecast files into a single 12-hour accumulation forecast.

    ***METv2.0/bin/pcp_combine ***
    *-add 20050807_060000.grb 6 *
    *20050807_120000.grb 6 *
    sample_fcst.nc -pcpdir data/2005080700

  - Summing 12 1-hourly accumulation observation files into a single 12-hour accumulated observation.

    ***METv2.0/bin/pcp_combine ***
    *-sum 00000000_000000.grb 1 *
    *20050807_120000.grb 12 *
    sample_obs.nc -pcpdir data/ST2ml

PCP-Combine: Example #1



Graphics produced using ncview

GRIB vs GRIB2

- **GRIB (or GRIB1): WRF postprocessor (WPP) produces GRIB1 format using copyGB**
  - Performs horizontal interpolation and destaggering (in the case of WRF-NMM) onto a defined grid.
  - Useful for both cores in creating an output grid not fixed by the model integration domain.
    - http://www.dtcenter.org/wrf-nmm/users/downloads/
- **GRIB2: NCEP and other WMO organizations have historical data in GRIB1 but now use GRIB 2 as the standard for gridded binary data.**
- **At least two GRIB2toGRIB1 converters available**
  - **NCEP cnvgrib**
    - http://www.nco.ncep.noaa.gov/pmb/codes/GRIB2/
  - **UCAR/CISL Grib Converter**
    - http://dss.ucar.edu/libraries/grib/c.html

Data Inventory Tools

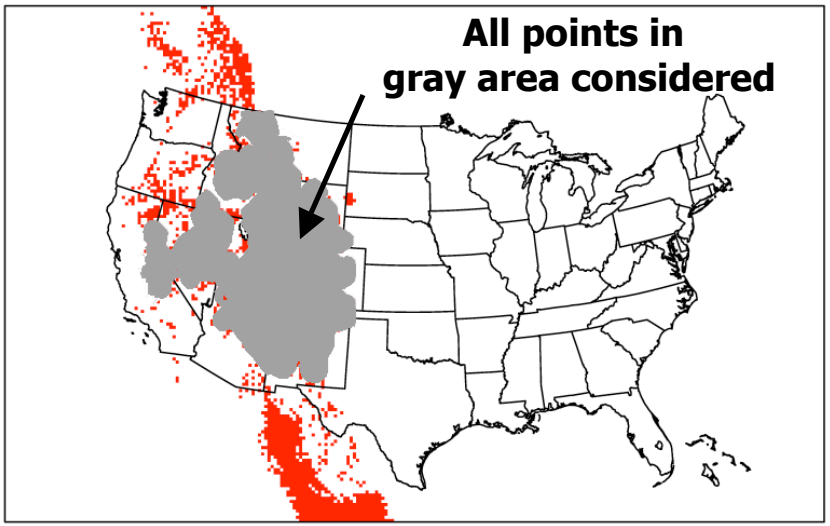
- **wgrib** – dumps GRIB1 headers and data.
  - http://www.cpc.ncep.noaa.gov/products/wesley/wgrib.html
- **wgrib2** – dumps GRIB2 headers and data.
  - http://www.cpc.ncep.noaa.gov/products/wesley/wgrib2/
- **ncdump -** dumps NetCDF headers and data.
  ncview – plots gridded NetCDF data.
  - http://www.unidata.ucar.edu/software/netcdf/
- **GrADS** – command line interface to produce plots.
  - http://www.iges.org/grads/downloads.html
- **NCL** – command line interface to produce plots.
  - http://www.ncl.ucar.edu/
- **IDV** – gui-driven visualization of many gridded and point datasets.
  - http://www.unidata.ucar.edu/software/idv/

4. Gen Poly Mask Tool

- **Stands for "General Polyline Mask"**
- **Functionality:**
  - Pre-computes a lat/lon polyline to generate a 0/1 mask field to be applied to your data.
  - Applies this mask once – prior to running Point-Stat or Grid-Stat
  - No configuration file.
- **Data formats:**
  - Reads Ascii formatted polyline file.
  - Reads GRIB file.
  - Reads NetCDF files from PCP-Combine.
  - Writes gridded NetCDF file of 0/1 mask.

Example polyline mask

Coordinates generated from underlying points of Elevation > 5000 ft.



Example polyline file:
data/poly/CONUS.poly

```
CONUS
35.81352 -122.33439
36.01565 -121.51420
35.35092 -121.27343
34.68935 -121.03940
34.03100 -120.81184
34.21258 -120.01634
33.55392 -119.80508
...
```