

File Formats and Pre-Processing

- File Formats
- Pre-processing Tools
- Useful Links

FILE FORMATS

Supported File Formats

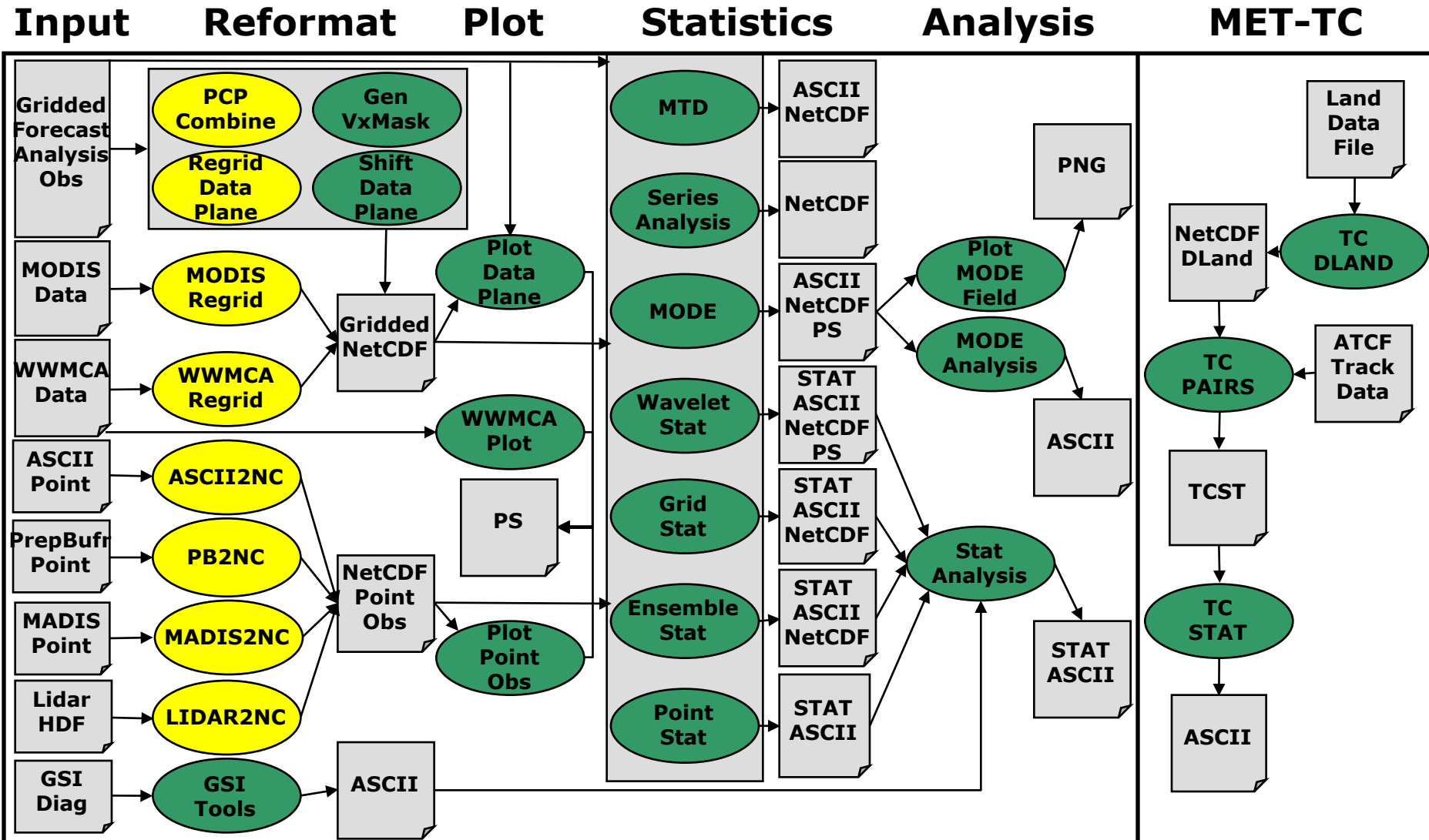
- **Forecasts**
 - **GRIB1** – GRIB1 Binary file
 - **GRIB2** – GRIB version 2 disabled by default (*--enable-grib2*)
 - **NetCDF** – Output from wrf_interp WRF-ARW utility, CF-Compliant versions 3 and 4, and internal MET NetCDF format
- **Gridded Analyses**
 - Same as Forecast file formats
 - GRIB Stage II/IV, MRMS, URMA, Model Analyses
 - **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/>
- **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/>

PRE-PROCESSING TOOLS

Pre-Processing / Reformatting



Data Reformating 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.

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

- **met-8.0/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
...

```

What obs are in a PREPBUFR file?

**>less **
ndas.t00z.prepbufr.tm12.2007
0401.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.

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

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. <i>AccPrecip</i> or <i>61</i> , <i>MSLP</i> or <i>2</i> , <i>Temp</i> or <i>11</i> , 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

- `met-8.0/bin/ascii2nc sample_obs.txt
sample_ascii.nc -v 2`

<pre>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" ; ... ;</pre>	<p>Result of ncdump -h</p> <p>Result of ncdump -v obs_arr</p>	<pre>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, ...</pre>
---	---	--

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, ROAB, 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]`

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

MADIS2NC: Run

- `met-8.0/bin/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,
```

4. Regrid_Data_Plane Tool

- **Functionality:**
 - Stand-alone tool implementing the automated regridding capability of the MET statistics tools.
 - Extract one or more user-specified fields from the input data file.
 - Regrid to the output grid using the specified interpolation method and width.
 - No configuration file.
- **Data formats:**
 - Reads any MET supported gridded data file (i.e. GRIB1/2 and flavors of NetCDF).
 - Writes gridded NetCDF as input to the MET statistics tools.

Regrid-Data-Plane: Usage

Usage: regrid_data_plane

input_filename

to_grid

output_filename

-field string

[-method type]

[-width n]

[-shape type]

[-vld_thresh n]

[-name list]

[-log file]

[-v level]

[-compress level]

input_filename	Input gridded data file name
to_grid	Output grid as a named grid, gridded data file, or grid specification
output_filename	Output NetCDF file name
-field string	Input field configuration string (may be used multiple times)
-method type	Interpolation method
-shape type	Interpolation shape (SQUARE or CIRCLE)
-width n	Interpolation width
-vld_thresh n	Interpolation required valid data ratio
-name list	Output NetCDF variable name(s)

Regrid-Data-Plane: Run

- **met-8.0/bin/regrid_data_plane **
in.grb G212 tmp_p500_G212.nc \
-field 'name="TMP"; level="P500";'
- **met-8.0/bin/regrid_data_plane **
in.grb gfs.t06z.pgrb2full.0p50.f078 \
surface_winds.nc \
-field 'name="UGRD"; level="Z10";' \
-field 'name="VGRD"; level="Z10";' \
-field 'name="WIND"; level="Z10";' \
-name UWind,VWind,WindSpeed

5. 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 GRIB1, GRIB2, or pinterp or CF compliant NetCDF 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]`

`[-field string]`

`[-name variable_name]`

`[-log file]`

`[-v level]`

`[-compress level]`

<code>-sum</code>	Accumulates data over multiple files. <i>Sum_args</i>: (init_time, in_accum, valid_time, out_accum, out_file, -pcpdir path, -pcprx reg_exp)
<code>-add</code>	Accumulates data over two files. <i>Add_args</i>: (in_file1, Accum1, in_file2, Accum2, out_file).
<code>-subtract</code>	Subtracts data over two files. <i>Sub_args</i>: (in_file1, Accum1, in_file2, Accum2, out_file).
<code>-field</code>	Defines the data to be extracted from the input files.
<code>-name</code>	Name of combined variable in output NetCDF file.

PCP-Combine: Sum

- **Two examples of the `-sum` option**

- 1) Sum two 6-hourly accumulation forecast files into a single 12-hour accumulation forecast.

```
met-8.0/bin/pcp_combine \  
-sum 20050807_000000 6 20050807_120000 12  
sample_fcst.nc -pcpdir data/2005080700
```

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

```
met-8.0/bin/pcp_combine \  
-sum 00000000_000000 1 \  
20050807_120000 12 \  
sample_obs.nc -pcpdir data/ST2ml
```

PCP-Combine: Add and Subtract

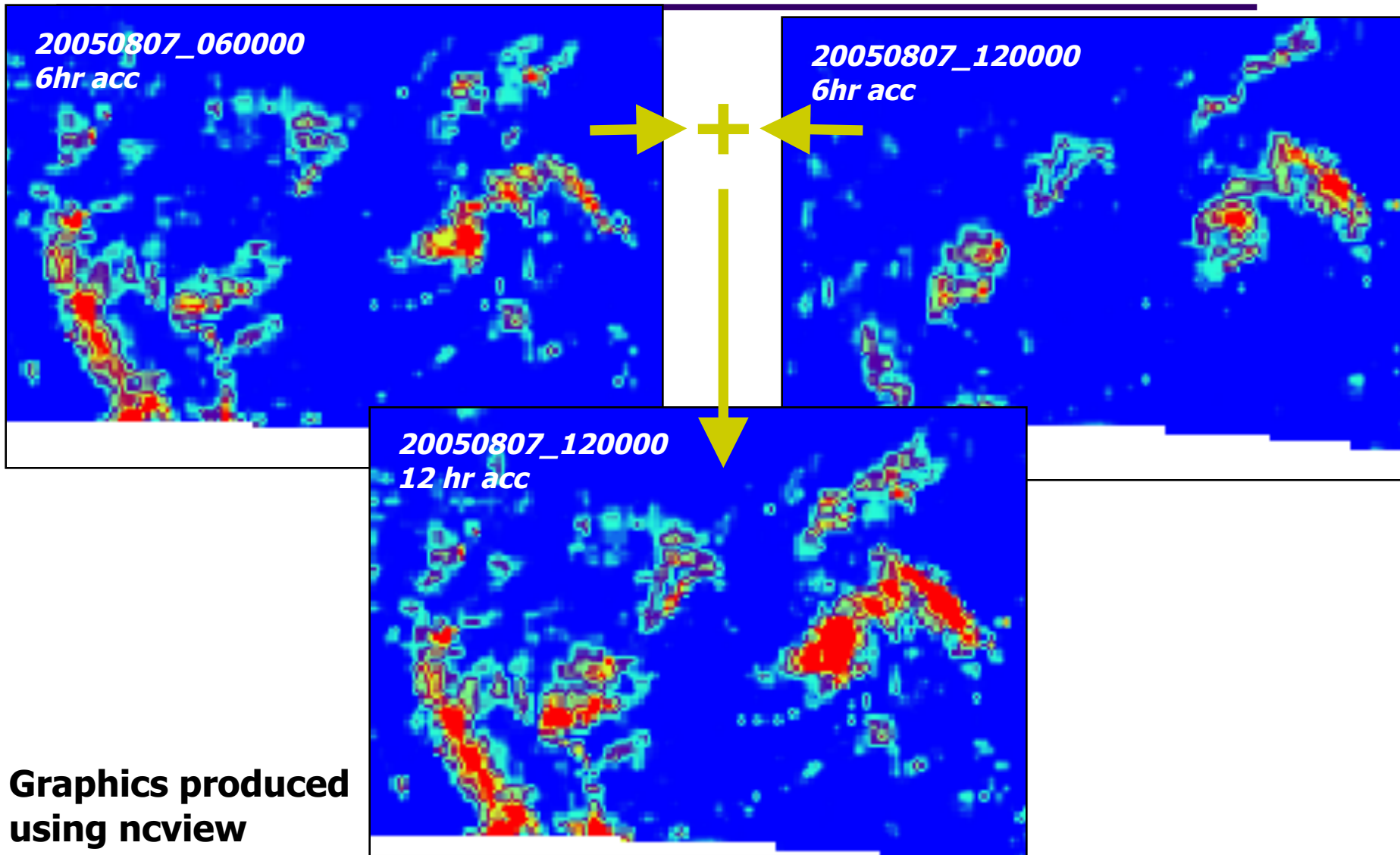
- Use **-add** option for already binned precipitation:
 - Adding two 6-hourly accumulation forecast files into a single 12-hour accumulation forecast.

```
met-8.0/bin/pcp_combine -add \  
20050807_060000.grb 6 \  
20050807_120000.grb 6 \  
APCP_12_20050807_120000.nc
```

- Use **-subtract** option for “runtime” accumulations:
 - Subtract 36 hour accumulation minus 12 hour accumulation for 24 hours in between.

```
met-8.0/bin/pcp_combine -subtract \  
nam_2012040900_F036.grib 36 \  
nam_2012040900_F012.grib 12 \  
nam_2012040900_F036_APCP_24.nc
```

PCP-Combine: Example #1



Graphics produced
using ncview

SPECIALIZED SATELLITE PRE-PROCESSING TOOLS

6. MODIS-Regrid Tool

- **Depends on HDF4/HDFEOS libraries.**
- **Compilation disabled by default (*--enable-modis*)**
- **Functionality:**
 - Reformat MODIS satellite observations into intermediate NetCDF format.
 - No configuration file.
- **Data formats:**
 - Reads MODIS level 2 data.
 - Writes gridded NetCDF as input to the MET statistics tools.

MODIS-Regrid: Usage

Usage: modis_regrid

-data_file path

-field name

-out path

-scale value

-offset value

-fill value

[-units text]

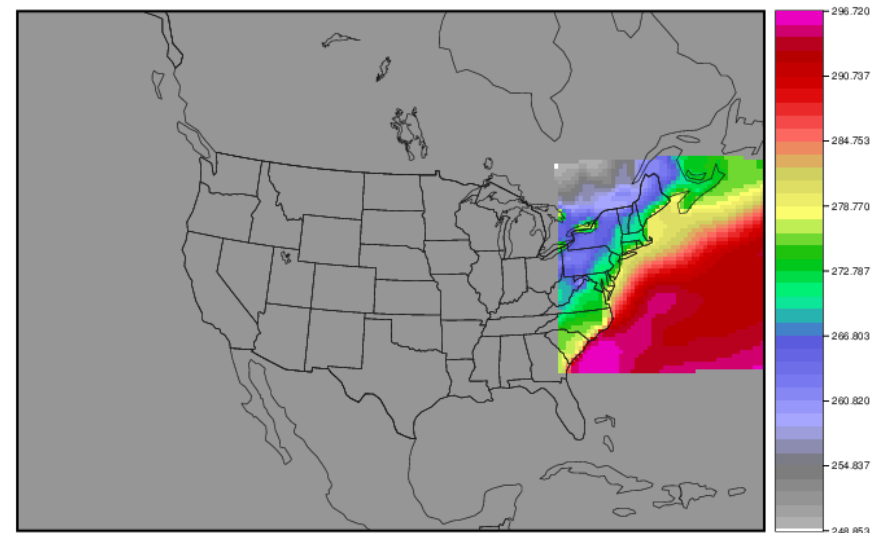
[-compress level]

modis_file

-data_file path	Gridded data file defining output grid
-field name	Field to process in MODIS file, e.g. temperature
-out path	Output NetCDF file name
-scale value	Scale factor to use
-offset value	Offset factor
-fill value	Bad data value
-units text	Units string to be written to the output file
modis_file	Input file MODIS file name

MODIS-Regrid: Run

- `met-8.0/bin/modis_regrid -field Cloud_Fraction \`
`-data_file grid_file -out t2.nc \`
`-units percent -scale 0.01 -offset 0 -fill 127 \`
`~/modis_regrid_test_data/modisfile`



m_sfc_temp.nc

7. WWMCA-Regrid Tool

- **Functionality:**
 - Reformat Air Force binary World Wide Merged Cloud Analysis into intermediate NetCDF format.
 - No configuration file.
- **Data formats:**
 - Reads binary WWMCA files.
 - Writes gridded NetCDF as input to the MET statistics tools.

WWMCA-Regrid: Usage

Usage: wwmca_regrid

-out filename

-config filename

-nh filename

[pt_filename]

-sh filename

[pt_filename]

[-log file]

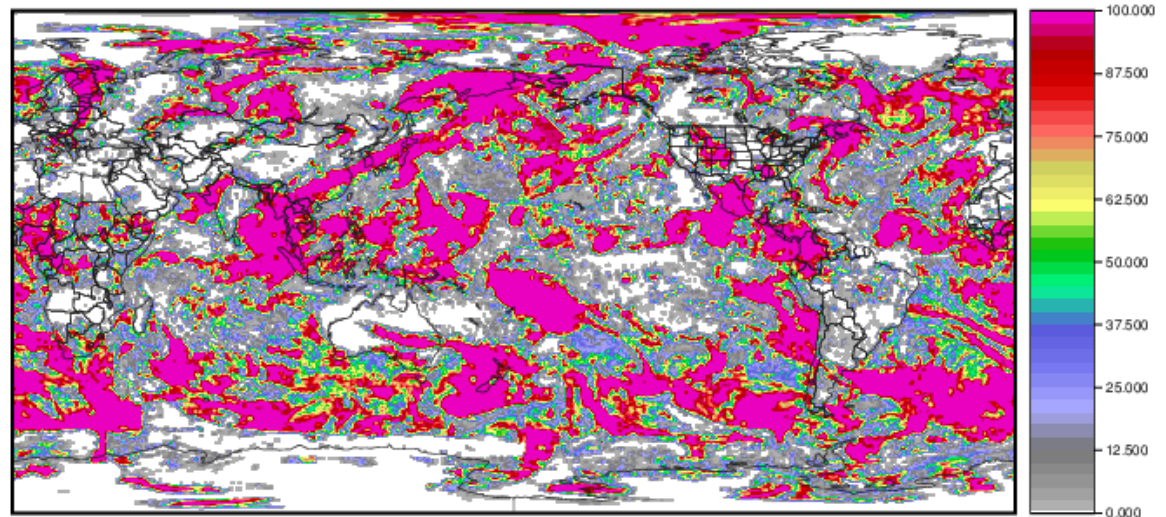
[-v level]

[-compress level]

-out filename	Output NetCDF file name
-config filename	Configuration file name
-nh filename	Northern Hemisphere data file
-sh filename	Southern Hemisphere data file
[pt_filename]	Pixel time files for the Northern and Southern hemispheres to mask data by pixel age

WWMCA-Regrid: Run

- `met-8.0/bin/wwmca_regrid \`
 `-config WWMCARegridConfig \`
 `-nh WWMCA_TOTAL_CLOUD_PCT_NH_2009083005 \`
 `-sh WWMCA_TOTAL_CLOUD_PCT_SH_2009083005 \`
 `-out WWMCA_TOTAL_CLOUD_PCT_2009083005_GFS_LATLON.nc \`
 `-v 2`



WWMCA_TOTAL_CLOUD_PCT_2009083005_GFS_LATLON.nc

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

- `met-8.0/bin/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
```

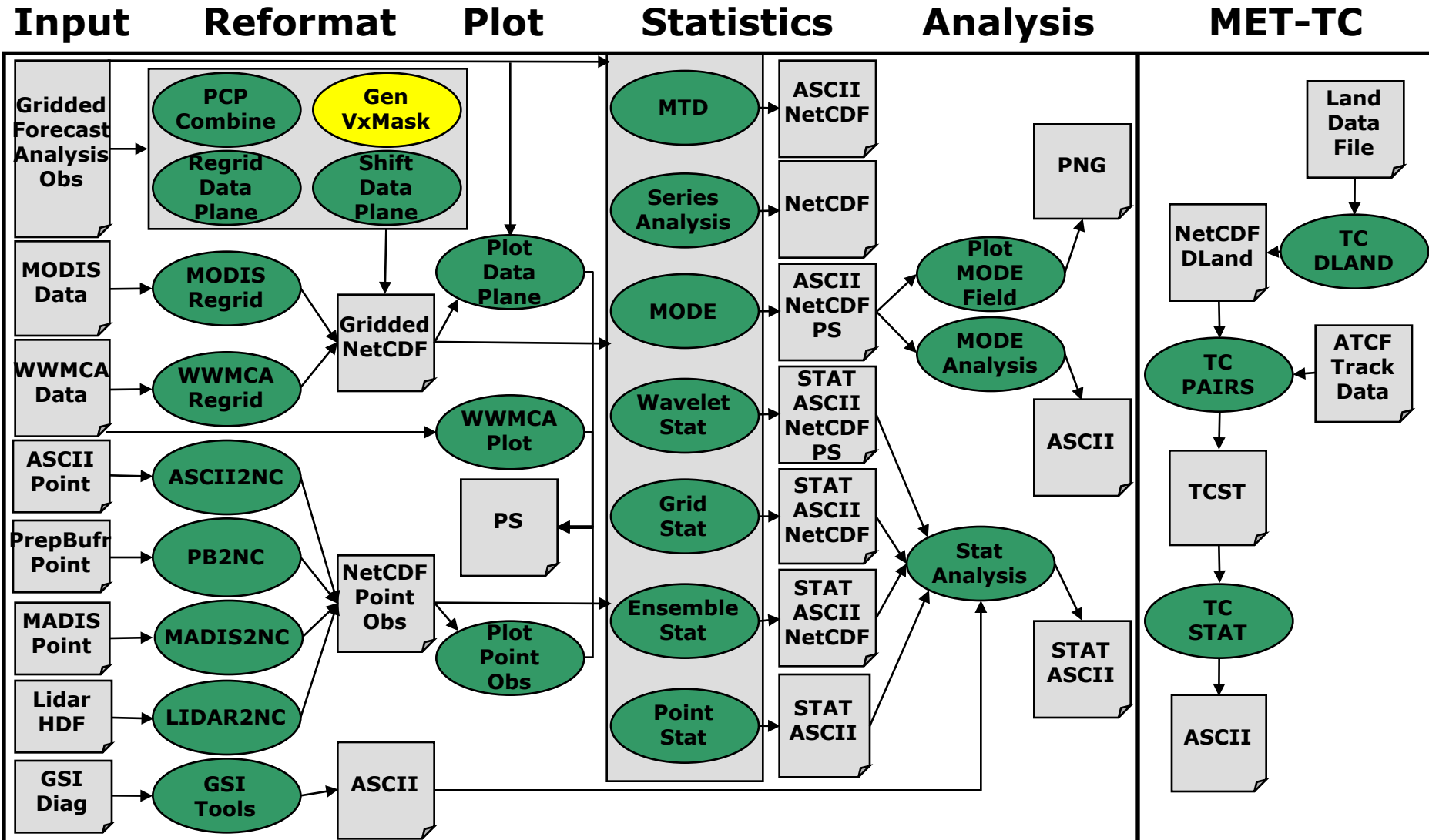
```
obs_arr =
0, 500, _, 0, 1,
0, 501, 995.6906, 142.5747, 142.5747,
0, 502, 865.9296, 1160.472, 1160.472,
0, 503, 995.6906, 142.5747, 0,
0, 504, 995.6906, 142.5747, 100,
0, 601, 995.6906, 142.5747, 2,
0, 602, 995.6906, 142.5747, 0,
0, 603, 995.6906, 142.5747, 0,
0, 604, 995.6906, 142.5747, 3,
0, 600,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

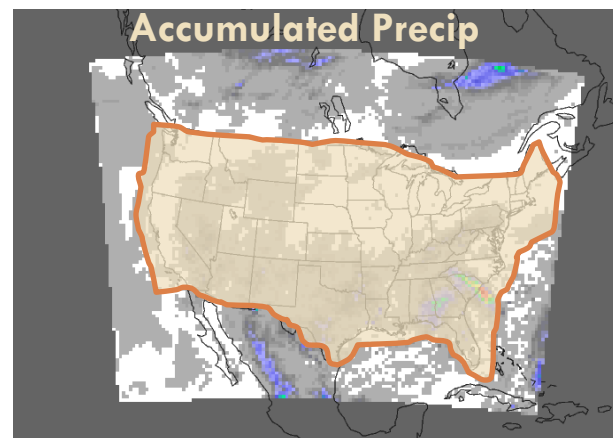
Gen-Vx-Mask Tool

Gen-Vx-Mask Tool

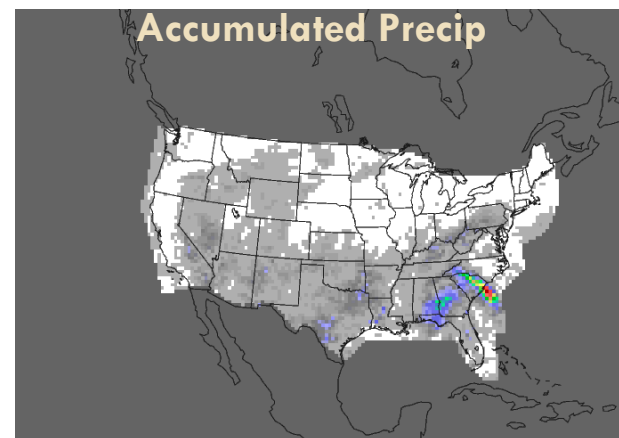
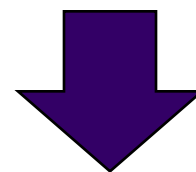


Gen-Vx-Mask: Overview

- **Generate Verification Mask**
 - Replaces earlier Gen-Poly-Mask and Gen-Circle-Mask tools
- **Purpose:**
 - Generate mask once for a domain and use the output many times.
- **Functionality:**
 - Generate a 0/1 bitmap mask field to define which grid points are included in statistics.
 - Support multiple masking methods.
 - Run iteratively to define complex masking region.
 - Define mask once prior to running the MET statistics tools.
 - No configuration file.
- **Data formats:**
 - Reads gridded data files.
 - Reads ASCII formatted lat/lon file.
 - Writes gridded output NetCDF mask file.



CONUS	
31.1931	-120.4211
31.2291	-120.4976
31.2650	-120.5741
31.3009	-120.6123
31.3369	-120.6506
31.3728	-120.6888
31.4087	-120.6888
31.4447	-120.7270
992 more points...	



Gen-Vx-Mask: Usage

Usage: gen_vx_mask

input_file

mask_file

out_file

[-type string]

[-input_field string]

[-mask_field string]

[-complement][-union]

[-intersection][-symdiff]

[-thresh string]

[-height n][-width n]

[-value n]

[-name string]


[-log file]

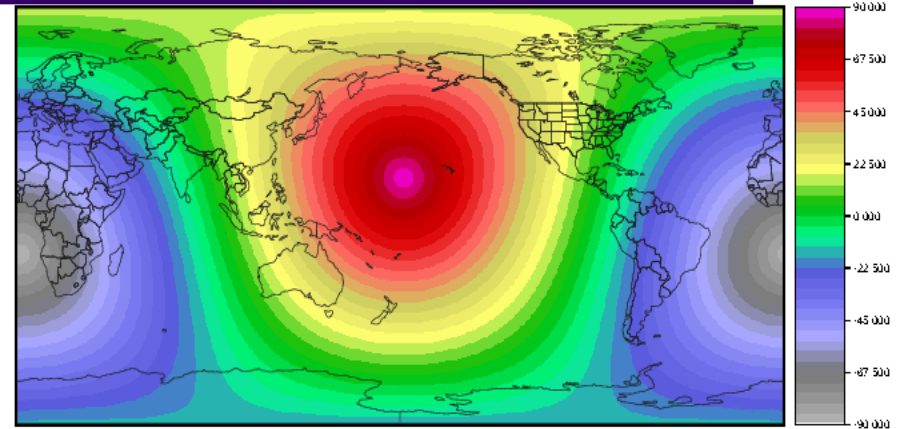
[-v level]

[-compress level]

input_file	Defines grid for the mask
mask_file	Defines the spatial masking area
out_file	Output NetCDF file name
-type string	poly, box, circle, track, grid, data, solar_alt, solar_azimuth, or shape
-input_field	Field for initial value at each grid point (instead of 0)
-mask_field	Field for data masking
-complement	Define complement of the mask
-union -intersection -symdiff	Control logic for combining -input_field and current mask
-thresh	Threshold for circle, track, data, solar_alt, and solar_azimuth types
-height, -width, -shapeno	Height and width for box type. Shapefile number.
-value, -name	Output mask value and variable name

Gen-Vx-Mask: Types

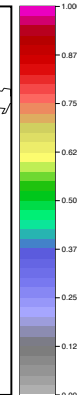
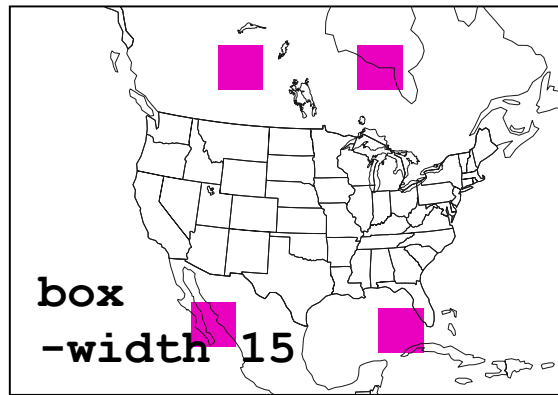
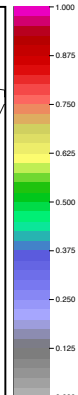
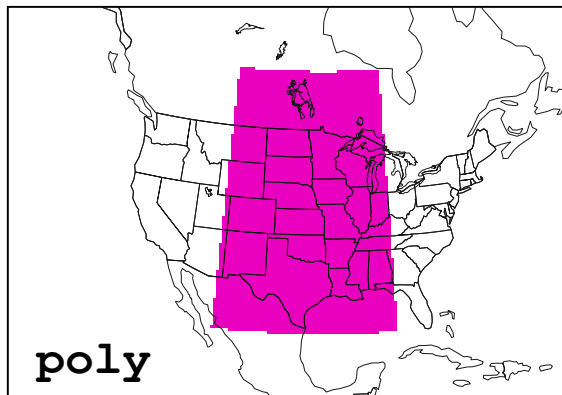
- mask_file = Lat/Lon file
 - Polyline (poly)
 - Box
 - Circle
 - Track
- mask_file = gridded data file
 - Grid
 - Data
 - Lat or Lon
- mask_file = gridded data file or timestamp
 - Solar Altitude (solar_alt) 
 - Solar Azimuth (solar_azi)
- mask_file = shapefile
 - Shape



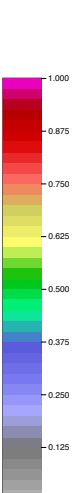
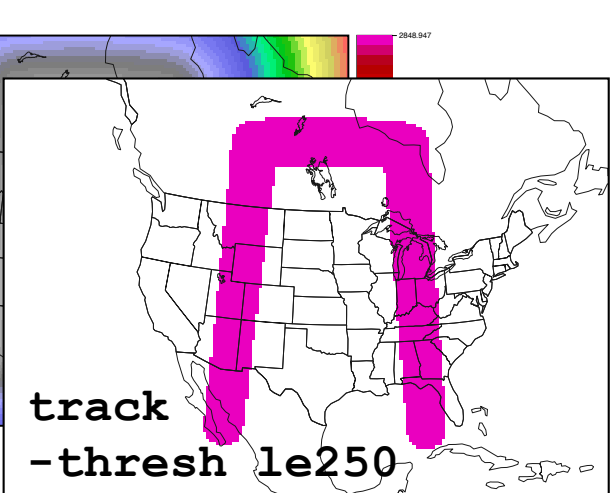
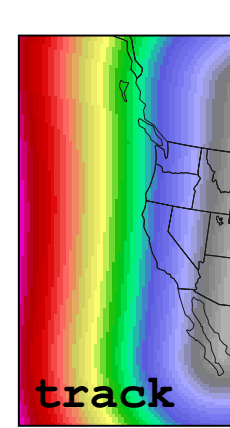
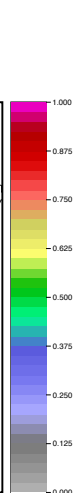
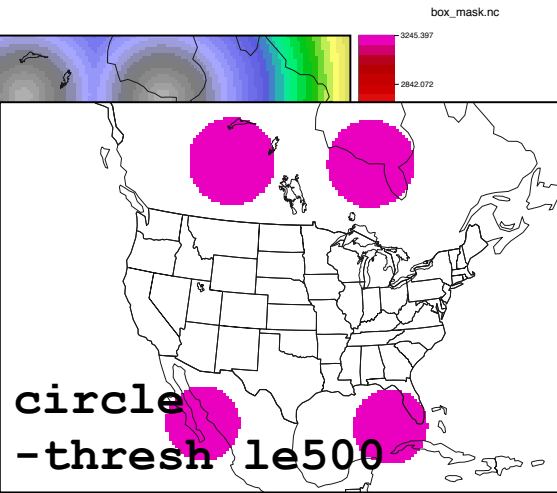
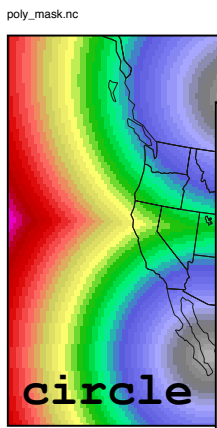
solar_alt_00.nc

Gen-Vx-Mask: Lat/Lon Types

```
gen_vx_mask wrfprs_ruc13_12.tm00 MyLatLonPoints.txt \
poly_mask.nc -type poly
```



MyLatLonPoints	
25	-110
55	-110
55	-85
25	-85



circle_mask_no_thresh.nc

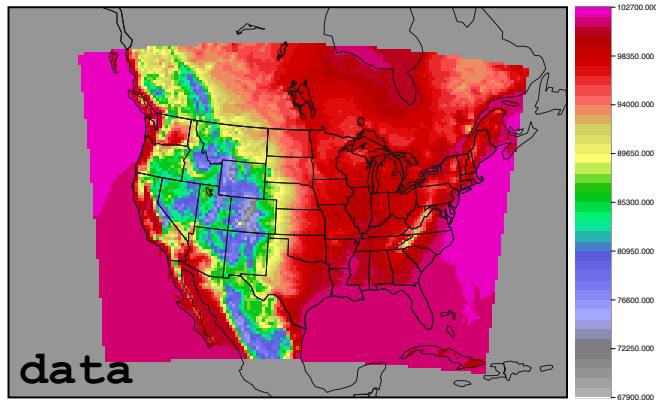
circle_mask_with_thresh.nc

track_mask_no_thresh.nc

track_mask_with_thresh.nc

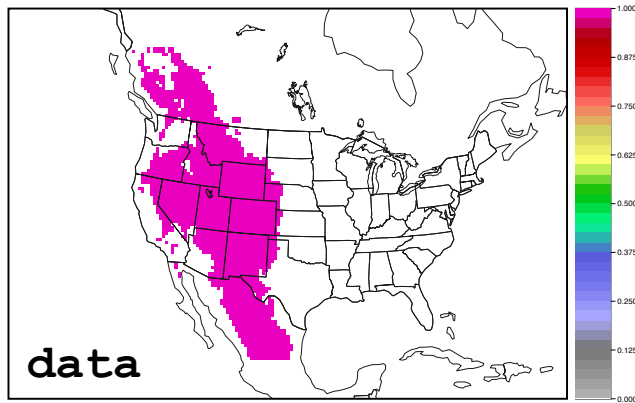
Gen-Vx-Mask: Data File Types

```
gen_vx_mask wrfprs_ruc13_12.tm00 d01_2009123112_02400.grib \
  grid_mask.nc -type grid
```

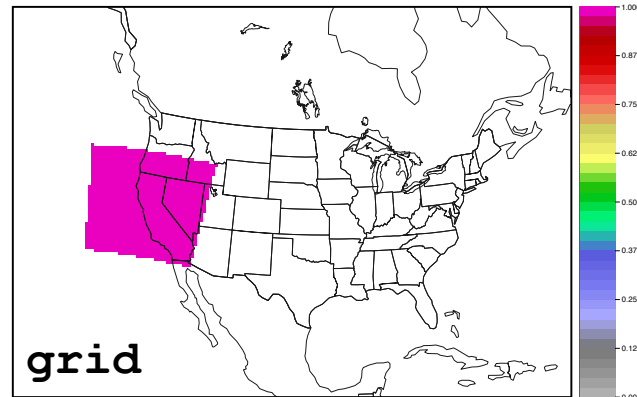


`-mask_field`

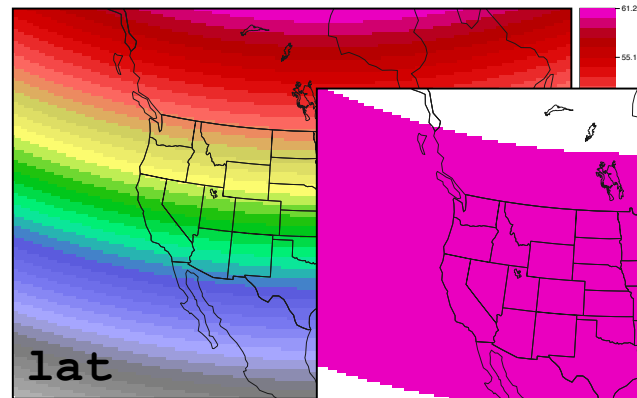
```
'name="PRES"; level="L0";'
```



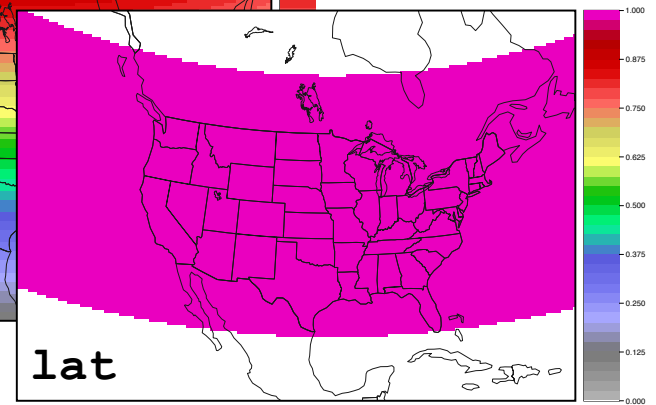
`-thresh 1e90000`



`grid_mask.nc`



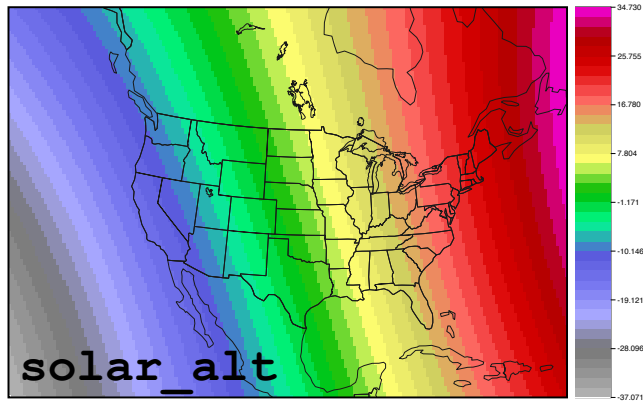
`lat_mask_no_thresh.nc`



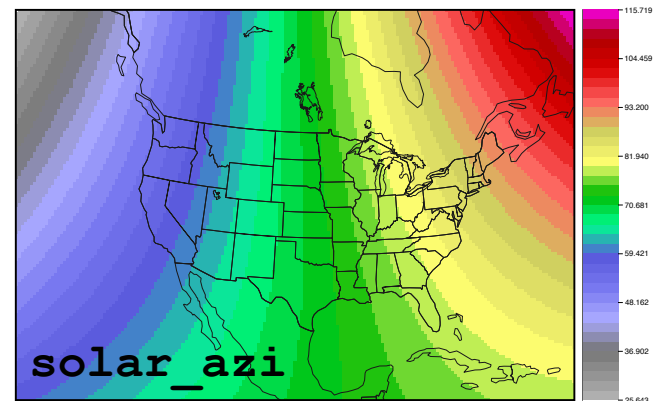
`-thresh 'ge25&&1e55'`

Gen-Vx-Mask: Timestamp Types

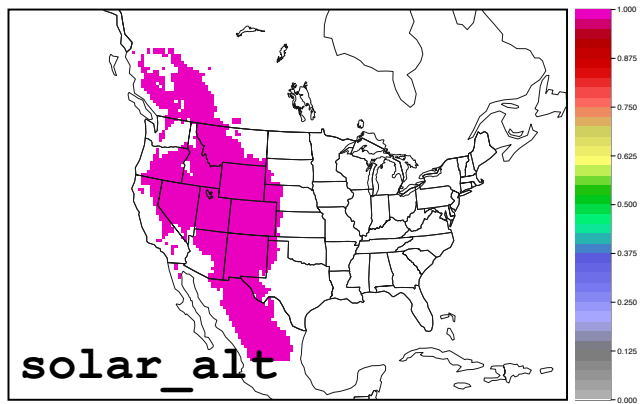
```
gen_vx_mask wrfprs_ruc13_12.tm00 20050807_12 \  
solar_alt_mask.nc -type solar_alt
```



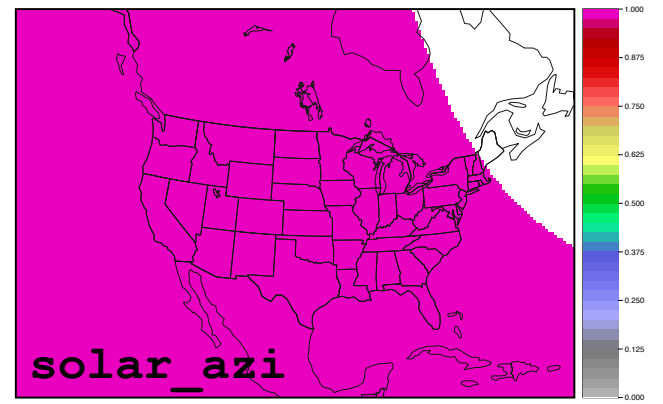
solar_alt_mask_no_thresh.nc



solar_azi_mask_no_thresh.nc



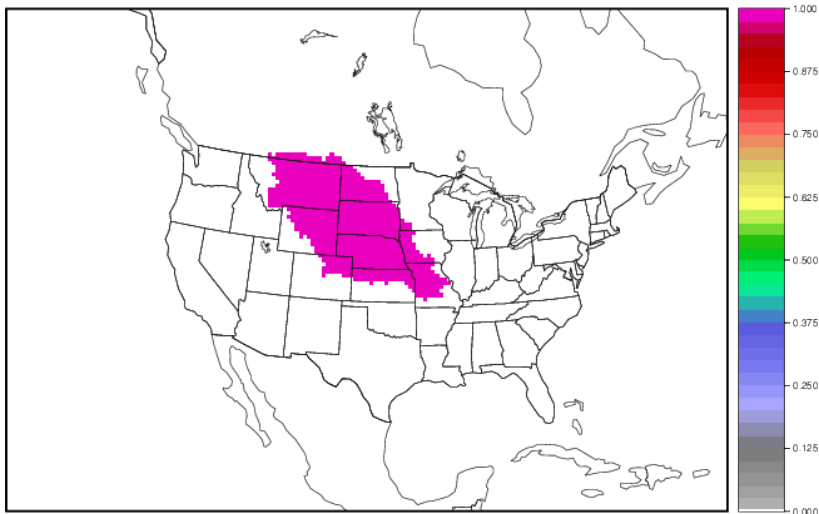
-thresh gt0
data_mask_wrfprs_ruc13_12.nc



-thresh 1e90
data_mask_wrfprs_ruc13_12.nc

Gen-Vx-Mask: Shapefiles

```
gen_vx_mask wrfprs_ruc13_12.tm00 rf12ja05.shp \  
G212_MBRFC.nc -type shape -shapeno 1 -name MBRFC
```

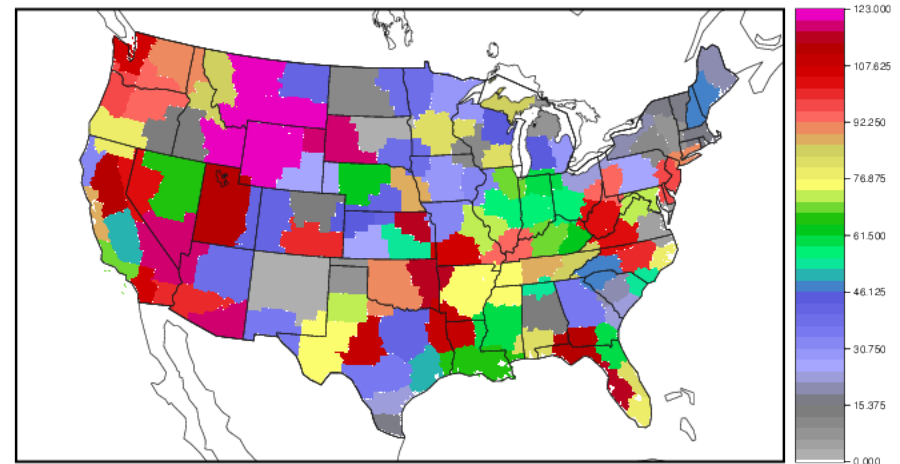


G212_MBRFC.nc

**Define Missouri Basin River
Forecast Center (MBRFC)
for NCEP Grid 212.**

**Run gen_vx_mask
iteratively to define NWS
County Warning Area for
HRRR domain.**

NWS County Warning Areas on HRRR Domain



hrrr_cwa_mask.nc

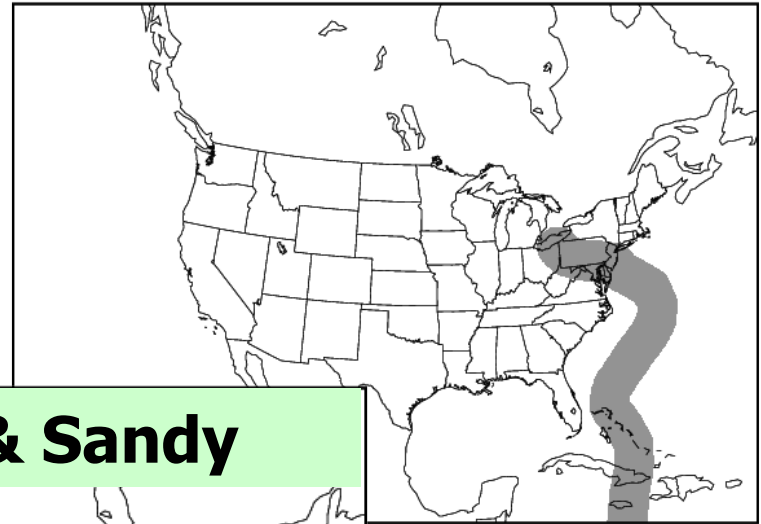
* Example from Dana Strom

Gen-Vx-Mask: Set Logic

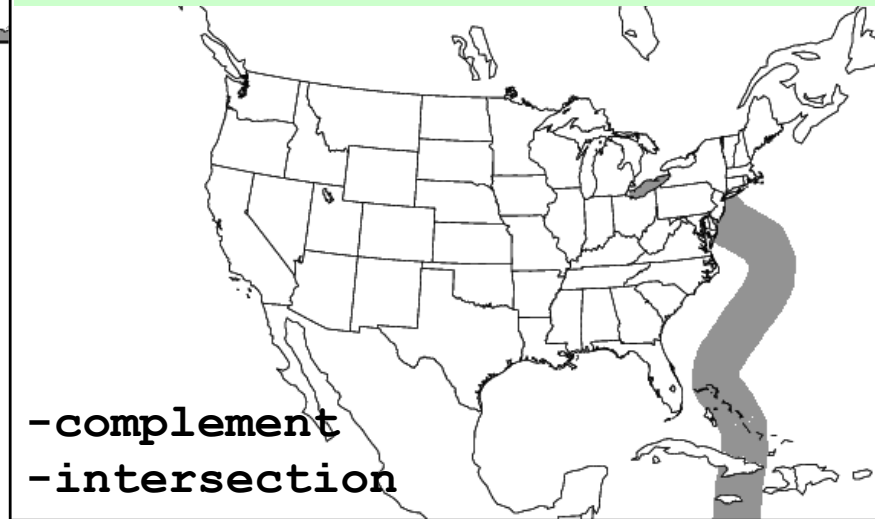
Land == 1



Sandy 200km

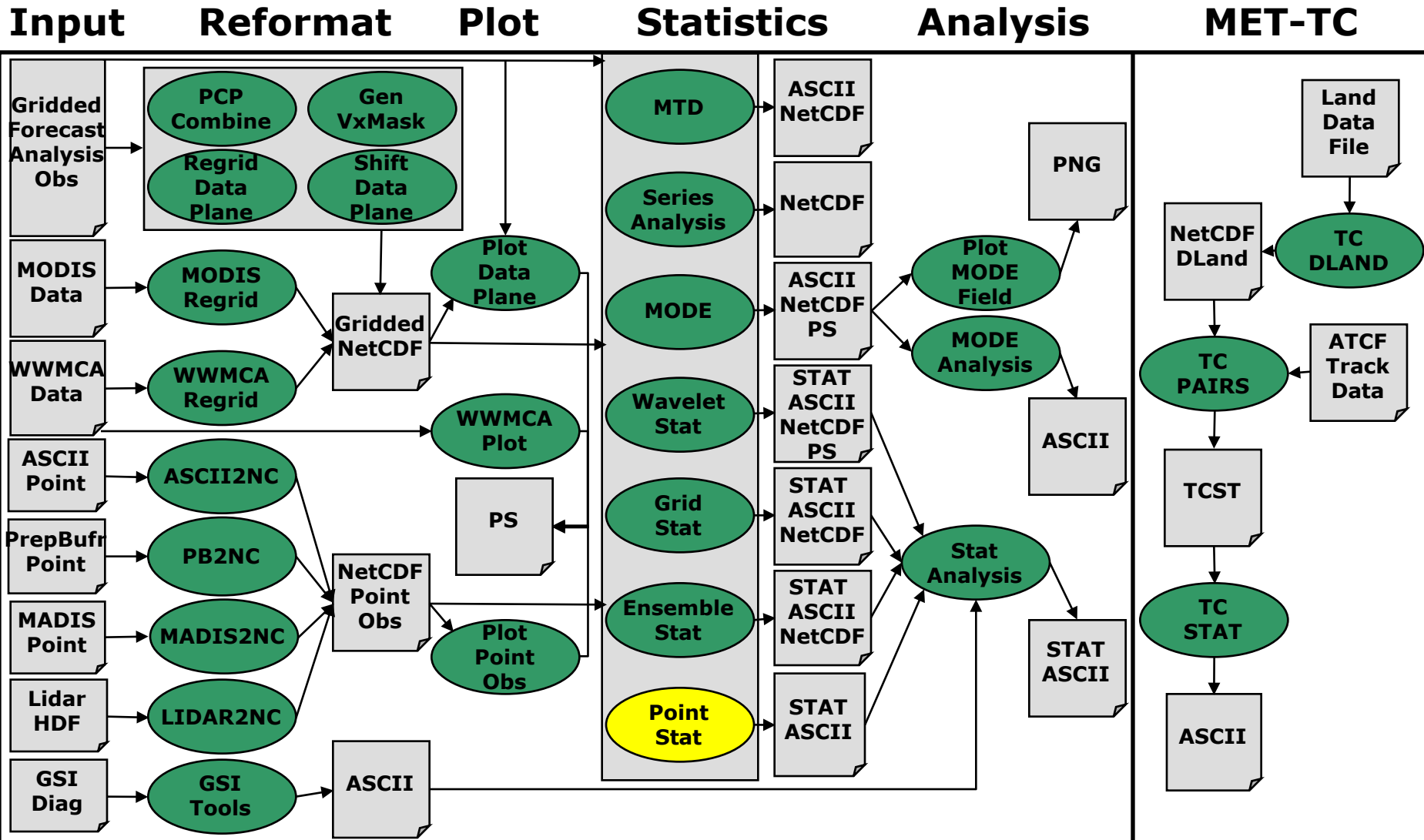


Not Land && Sandy

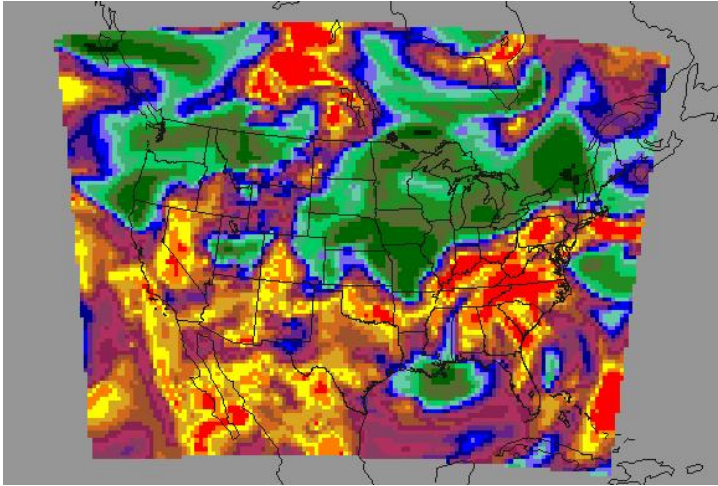


Point-Stat Tool

Point-Stat Tool

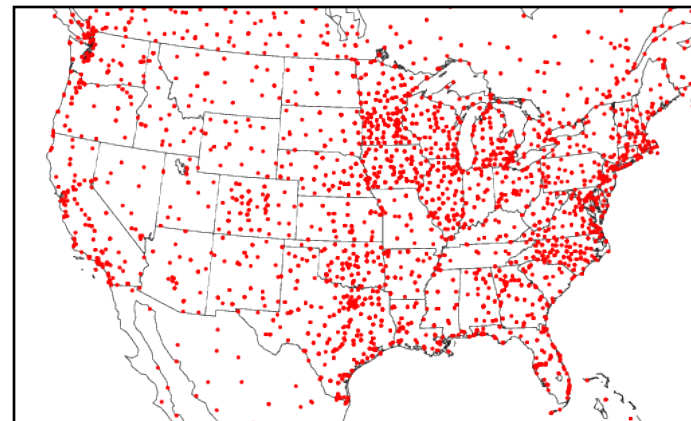


Point-Stat: Overview



- Compare **gridded forecasts** to **point observations**.
- Accumulate matched pairs over a defined area at a **single point in time**.
- Verify one or more variables/levels.
- Analysis tool provided to aggregate through time.

- **Verification methods:**
 - **Continuous** statistics for raw fields.
 - **Single and Multi-Category** counts and statistics for thresholded fields.
 - **Parametric and non-parametric confidence intervals** for statistics.
 - **Compute partial sums** for raw fields and/or the raw matched pair values.
 - **Methods for probabilistic** forecasts.
 - **HiRA spatial verification method**.



Point-Stat: Input/Output

- Input Files
 - Gridded forecast file
 - GRIB1 output of Unified Post-Processor (or other)
 - GRIB2 from NCEP (or other)
 - NetCDF from PCP-Combine, wrf_interp, or CF-compliant
 - Point observation file
 - NetCDF output of PB2NC, ASCII2NC, MADIS2NC, or LIDAR2NC
 - ASCII configuration file
- Output Files
 - ASCII statistics file with all output lines (end with “.stat”)
 - Optional ASCII files sorted by line type with a header row (ends with “_TYPE.txt”)

Point-Stat: Usage

Usage: point_stat

fcst_file

obs_file

config_file

[-point_obs netcdf_file]

[-obs_valid_beg time]

[-obs_valid_end time]

[-outdir path]

[-log file]

[-v level]

fcst_file	Gridded forecast file
obs_file	NC point observation file
config_file	ASCII configuration file
-point_obs	Additional NC point observation files
-obs_valid_beg	Beginning of valid time window for matching
-obs_valid_end	End of valid time window for matching
-outdir	Output directory to be used
-log	Optional log file
-v	Level of logging

Point-Stat: Configuration

- Many configurable parameters – only set a few:
 - 2-meter temperature.
 - Threshold temperatures near freezing.
 - Match to obs at the surface.
- Accumulate stats over all the points in the domain.
- Match observation to the nearest forecast value.
- Generate all output line types other than vector and probabilistic.

```
fcst = {
  message_type = [ "ADPSFC" ];
  field = [
    {
      name      = "TMP";
      level     = [ "Z2" ];
      cat_thresh = [ >273.0, >283.0, >293.0 ];
    }
  ];
};
obs = fcst;
```

```
mask = {
  grid = [ "FULL" ];
  poly = [];
  sid  = [];
  llpnt = [];
};
```

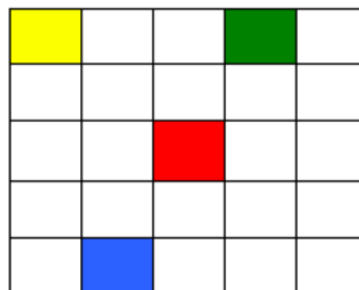
```
interp = {
  vld_thresh = 1.0;
  shape = SQUARE;
  type = [
    {
      method = UW_MEAN;
      width  = 1;
    }
  ];
};
```

```
output_flag = {
  fho  = BOTH;
  ctc  = BOTH;
  cts  = BOTH;
  mctc = BOTH;
  mcts = BOTH;
  cnt  = BOTH;
  sl112 = BOTH;
  sal112 = BOTH;
  vl112 = NONE;
  val112 = NONE;
  vcnt  = BOTH;
  pct   = NONE;
  pstd  = NONE;
  pjc   = NONE;
  prc   = NONE;
  ecnt  = NONE;
  eclv  = NONE;
  mpr   = BOTH;
};
```

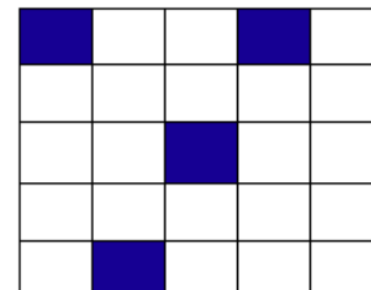
Point-Stat: HiRA Framework

- High Resolution Assessment (HiRA) verification logic is applied to deterministic forecasts matched to point observations.
- Process neighborhood values as...
 - an ensemble using neighborhood forecast values (ECNT line type).
 - a probability forecast (PCT, PSTD, PRC, and PJC line types).
- Allows for some spatial / temporal uncertainty by giving credit for being 'close'.
- Allows for comparison of deterministic and ensemble forecasts via the same set of probabilistic statistics.
- Also allows for comparison of models with different grid resolutions via adjustment of neighborhood size.

Model Forecast
White boxes = 0
Colored boxes > 0



Threshold Forecast
Blue boxes = event

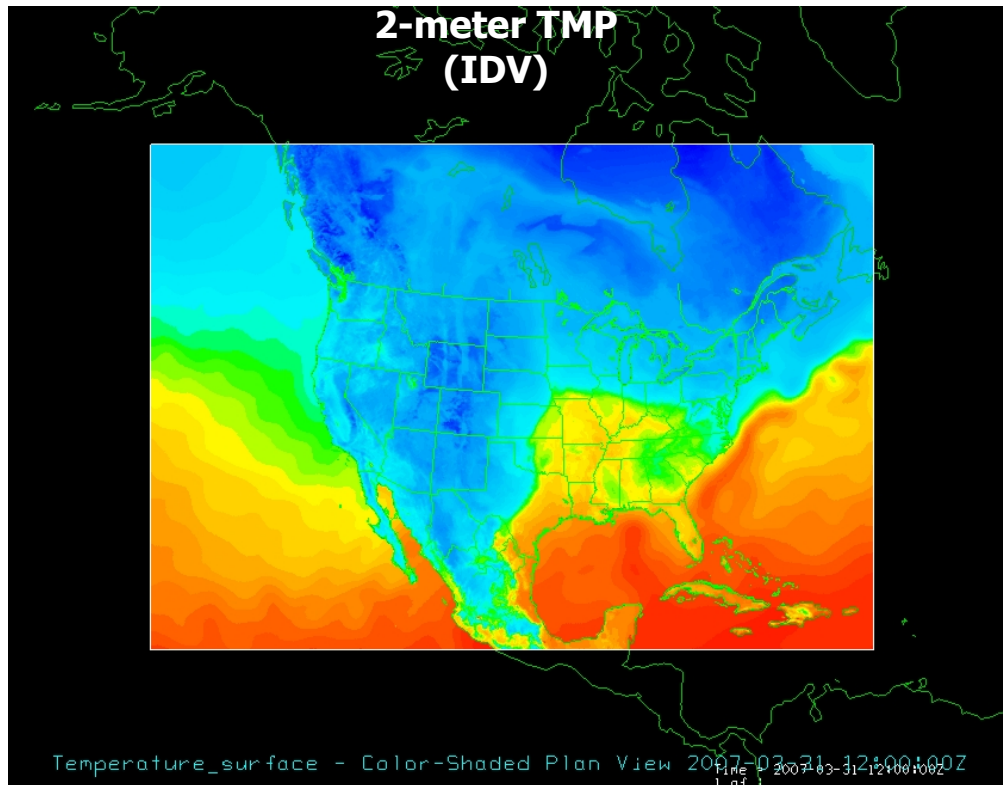


HiRA Ensemble:
Write ECNT using
neighborhood
ensemble.

HiRA Probabilities:
1x1 NBRHD = 1/1
3x3 NBRHD = 1/9
5x5 NBRHD = 4/25

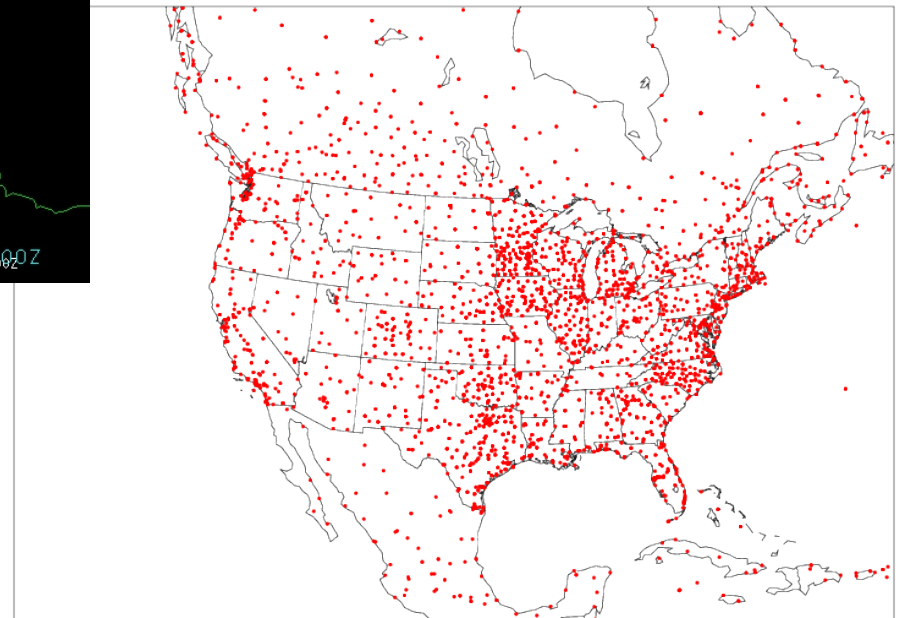
```
hira = {  
    // Enable or disable  
    flag      = TRUE;  
    // Neighborhood sizes (parity logic)  
    width     = [ 2, 3, 4, 5 ];  
    // Probability thresholds  
    cov_thresh = [ ==0.25 ];  
    vld_thresh = 1.0;  
    // Neighborhood shape  
    shape     = SQUARE;  
};
```

Point-Stat: Input



**4003 TMP ADPSFC Obs
(plot_point_obs)**

sample_pb.nc



Point-Stat: Run

- `met-8.0/bin/point_stat \`
`sample_fcst.grb sample_pb.nc \`
`PointStatConfig_TMPZ2 -outdir out -v 2`

```
DEBUG 1: Default Config File: met-8.0/share/met/data/config/PointStatConfig_default
DEBUG 1: User Config File: PointStatConfig_TMPZ2
DEBUG 1: Forecast File: sample_fcst.grb
DEBUG 1: Climatology File: none
DEBUG 1: Observation File: sample_pb.nc
DEBUG 2: -----
DEBUG 2: Reading data for TMP/Z2.
DEBUG 2: For TMP/Z2 found 1 forecast levels and 0 climatology levels.
DEBUG 2: -----
DEBUG 2: Searching 87752 observations from 9396 messages.
DEBUG 2: -----
DEBUG 2: Processing TMP/Z2 versus TMP/Z2, for observation type ADPSFC, over region FULL, for interpolation method
UW_MEAN(1), using 4003 pairs.
DEBUG 2: Computing Categorical Statistics.
DEBUG 2: Computing Multi-Category Statistics.
DEBUG 2: Computing Continuous Statistics.
DEBUG 2: Computing Scalar Partial Sums.
DEBUG 2: -----
DEBUG 1: Output file: out/point_stat_360000L_20070331_120000V.stat
DEBUG 1: Output file: out/point_stat_360000L_20070331_120000V_fho.txt
DEBUG 1: Output file: out/point_stat_360000L_20070331_120000V_ctc.txt
DEBUG 1: Output file: out/point_stat_360000L_20070331_120000V_cts.txt
DEBUG 1: Output file: out/point_stat_360000L_20070331_120000V_mctc.txt
DEBUG 1: Output file: out/point_stat_360000L_20070331_120000V_mcts.txt
DEBUG 1: Output file: out/point_stat_360000L_20070331_120000V_cnt.txt
DEBUG 1: Output file: out/point_stat_360000L_20070331_120000V_sl112.txt
DEBUG 1: Output file: out/point_stat_360000L_20070331_120000V_sal112.txt
DEBUG 1: Output file: out/point_stat_360000L_20070331_120000V_mpr.txt
```


Point-Stat: ASCII Output Types

- Statistics line types: 18 possible
 - Categorical – Single Threshold
 - Contingency table counts and stats (FHO, CTC, CTS, ECLV)
 - Categorical – Multiple Thresholds
 - NxN Contingency table counts and stats (MCTC, MCTS)
 - Scalars - raw fields
 - Continuous statistics (CNT) and partial sums (SL1L2, SAL1L2)
 - Vectors
 - Vector statistics (VCNT) and partial sums (VL1L2, VAL1L2)
 - Probabilistic
 - Nx2 Contingency table counts and stats (PCT, PSTD)
 - Continuous statistics and ROC curve (PJC, PRC)
 - Economic Cost/Loss value (ECLV)
 - HiRA (ECNT)
 - Raw matched pairs – a lot of data! (MPR)
- 22 header columns common to all line types
- Remaining columns specific to each line type

Point-Stat: Sample Output

1. **STAT** file output for sample run:
 - 1 line each for **CNT**, **SL1L2**, **MCTC**, **MCTS**
 - 3 lines each for **FHO**, **CTC**, **CTS**
 - 4,003 lines for **MPR**!
2. Additional **TXT** files for each line type

```
Output file: out/point_stat_360000L_20070331_120000V.stat
Output file: out/point_stat_360000L_20070331_120000V_fho.txt
Output file: out/point_stat_360000L_20070331_120000V_ctc.txt
Output file: out/point_stat_360000L_20070331_120000V_cts.txt
Output file: out/point_stat_360000L_20070331_120000V_mctc.txt
Output file: out/point_stat_360000L_20070331_120000V_mcts.txt
Output file: out/point_stat_360000L_20070331_120000V_cnt.txt
Output file: out/point_stat_360000L_20070331_120000V_sl1l2.txt
Output file: out/point_stat_360000L_20070331_120000V_sal1l2.txt
Output file: out/point_stat_360000L_20070331_120000V_mpr.txt
```

Point-Stat: CTC Output Line

VERSION	V8.0
MODEL	WRF
DESC	NA
FCST_LEAD	360000
FCST_VALID_BEG	20070331_120000
FCST_VALID_END	20070331_120000
OBS_LEAD	000000
OBS_VALID_BEG	20070331_103000
OBS_VALID_END	20070331_133000
FCST_VAR	TMP
FCST_LEV	Z2
OBS_VAR	TMP
OBS_LEV	Z2
OBTYPE	ADPSFC

VX_MASK	FULL
INTERP_MTHD	UW_MEAN
INTERP_PNTS	1
FCST_THRESH	>273.000
OBS_THRESH	>273.000
COV_THRESH	NA
ALPHA	NA
LINE_TYPE	CTC
TOTAL	4003
FY_OY (hits)	3111
FY_ON (f.a.)	78
FN_OY (miss)	215
FN_ON (c.n.)	599

Point-Stat: Matched Pairs

- Matched Pair (MPR) line type contains 1 line for each matched pair.
- Data overload!

TOTAL	INDEX	OBS_SID	OBS_LAT	OBS_LON	OBS_LVL	OBS_ELV	FCST	OBS	OBS_QC	...
4003	1	71600	43.93000	-60.01000	1010.79999	4.01053	272.00000	271.95001	NA	
4003	2	71616	46.43000	-71.93000	1016.09998	102.04903	268.00000	269.64999	NA	
4003	3	71629	44.23000	-78.36000	1004.50000	191.44466	273.00000	272.64999	NA	
4003	4	71028	51.67000	-124.40000	916.50000	872.82202	264.00000	265.25000	NA	
4003	5	71066	58.61000	-117.16000	973.90002	337.50449	272.00000	271.45001	NA	
4003	6	71104	52.18000	-122.04000	906.50000	938.08594	271.00000	264.64999	NA	
4003	7	71109	50.68000	-127.36000	1020.20001	22.03931	275.00000	275.64999	NA	
4003	8	71150	50.45000	-100.59000	949.09998	562.38477	272.00000	271.75000	NA	
4003	9	71177	57.13000	-61.47000	899.70001	834.87476	259.00000	254.64999	NA	
4003	10	71197	47.56000	-59.16000	1000.90002	40.06803	272.00000	269.95001	NA	
4003	11	71378	47.41000	-72.79000	1006.90002	169.37592	267.00000	266.14999	NA	
4003	12	71415	45.76000	-62.68000	1014.00000	1.99518	269.00000	269.04999	NA	
4003	13	71425	49.24000	-65.33000	1014.90002	28.96468	264.00000	267.25000	NA	
4003	14	71437	43.29000	-79.79000	1017.79999	77.03765	274.00000	276.35001	NA	
4003	15	71473	48.78000	-123.04000	1015.70001	23.93772	278.00000	280.95001	NA	
4003	16	71486	52.93000	-118.31000	896.20001	1021.55963	265.00000	270.35001	NA	
4003	17	71573	42.87000	-80.55000	996.09998	231.62808	273.00000	276.54999	NA	
4003	18	71579	49.89000	-97.13000	989.70001	230.44473	276.00000	274.35001	NA	
4003	19	71598	47.78000	-64.83000	1016.70001	5.00338	266.00000	267.54999	NA	
4003	20	71860	50.02000	-100.32000	874.20001	474.10379	273.00000	273.54999	NA	

Customizing Point-Stat and Grid-Stat Output

We'll restrict this discussion to an explanation of masking and interpolation.

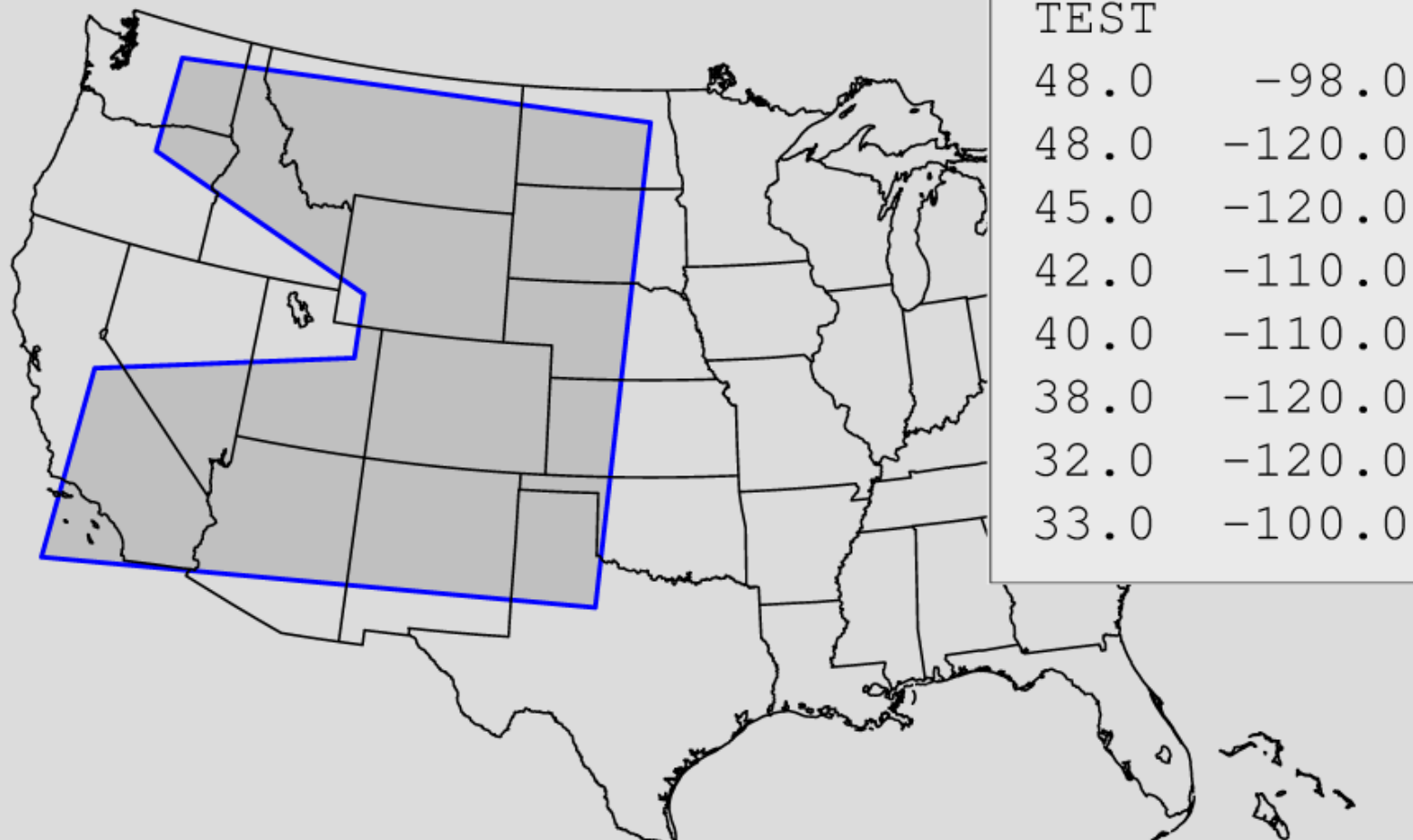
They're the parts that people have the most questions about, it seems.

What is **masking**? Many times you want to restrict your verification to some subregion. The subregion may be specified in several ways.

	Polyline	Grid	Stations	On/Off Bit Map
Point Stat	✓	✓	✓	✓
Grid Stat	✓	✓	N/A	✓

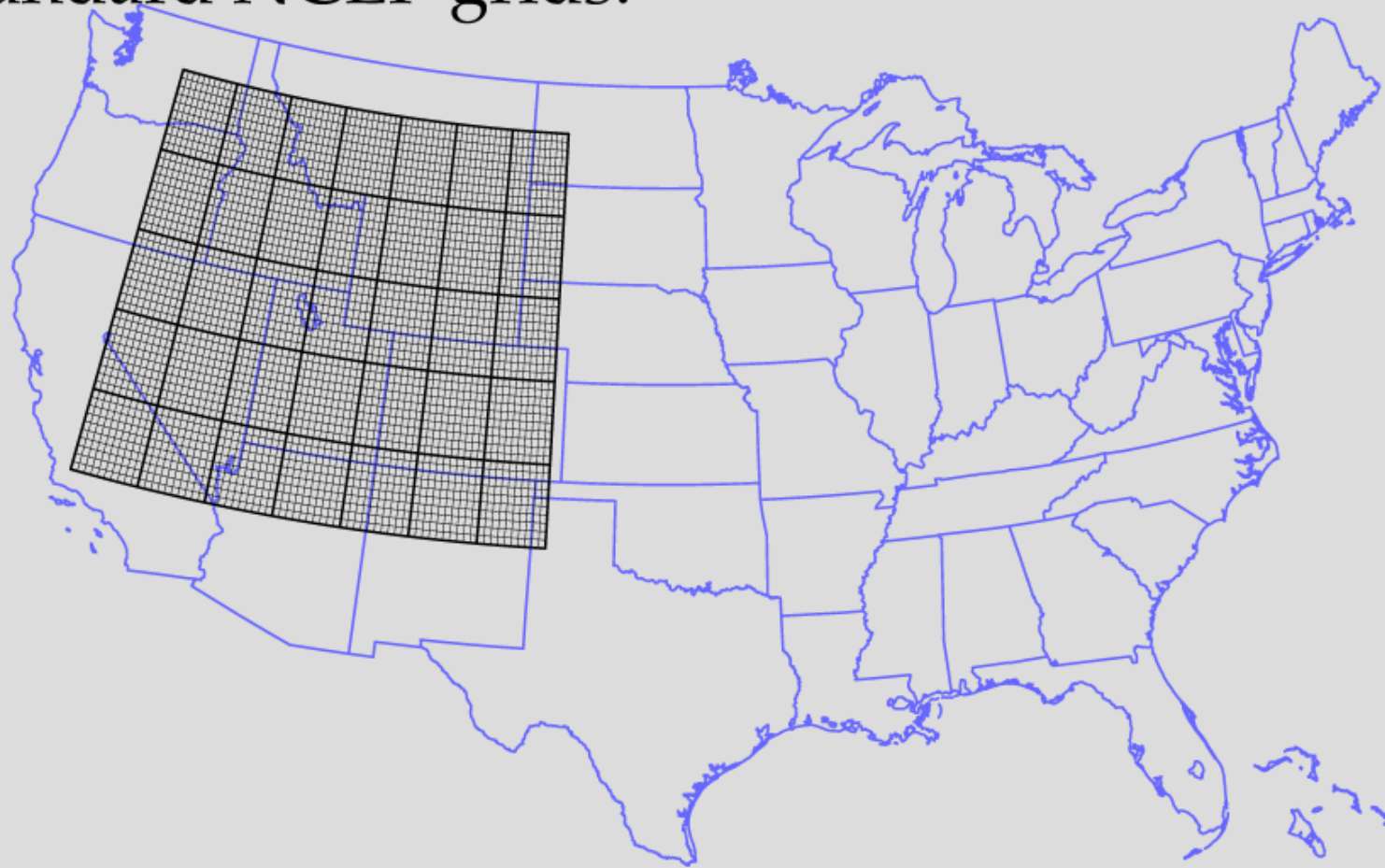
Masking by stations doesn't apply to Grid Stat.

Polyline masking: The user provides a closed lat/lon polyline boundary in a simple ASCII format.



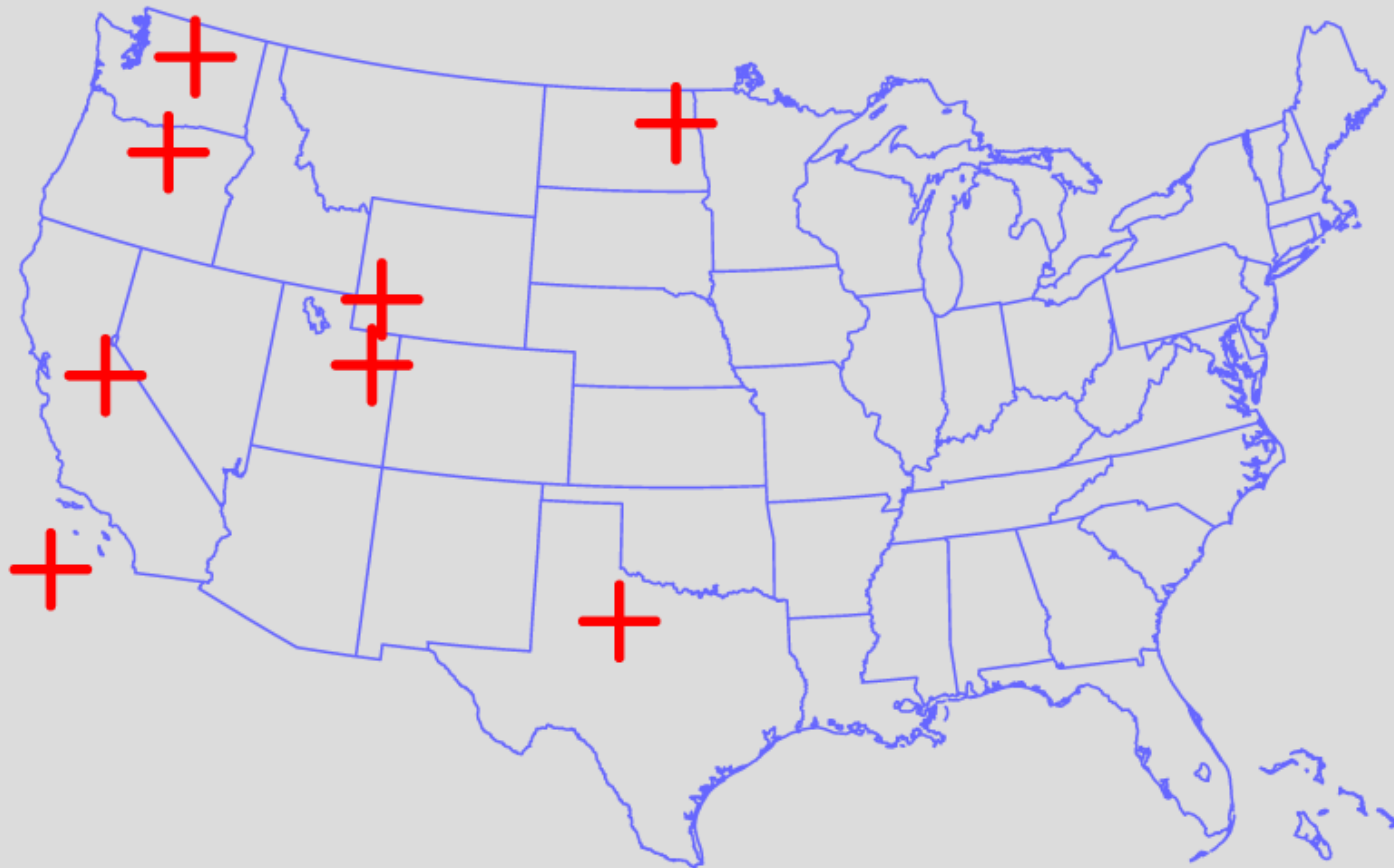
```
TEST
48.0   -98.0
48.0   -120.0
45.0   -120.0
42.0   -110.0
40.0   -110.0
38.0   -120.0
32.0   -120.0
33.0   -100.0
```

Grid Masking: The user can specify one of the standard NCEP grids.

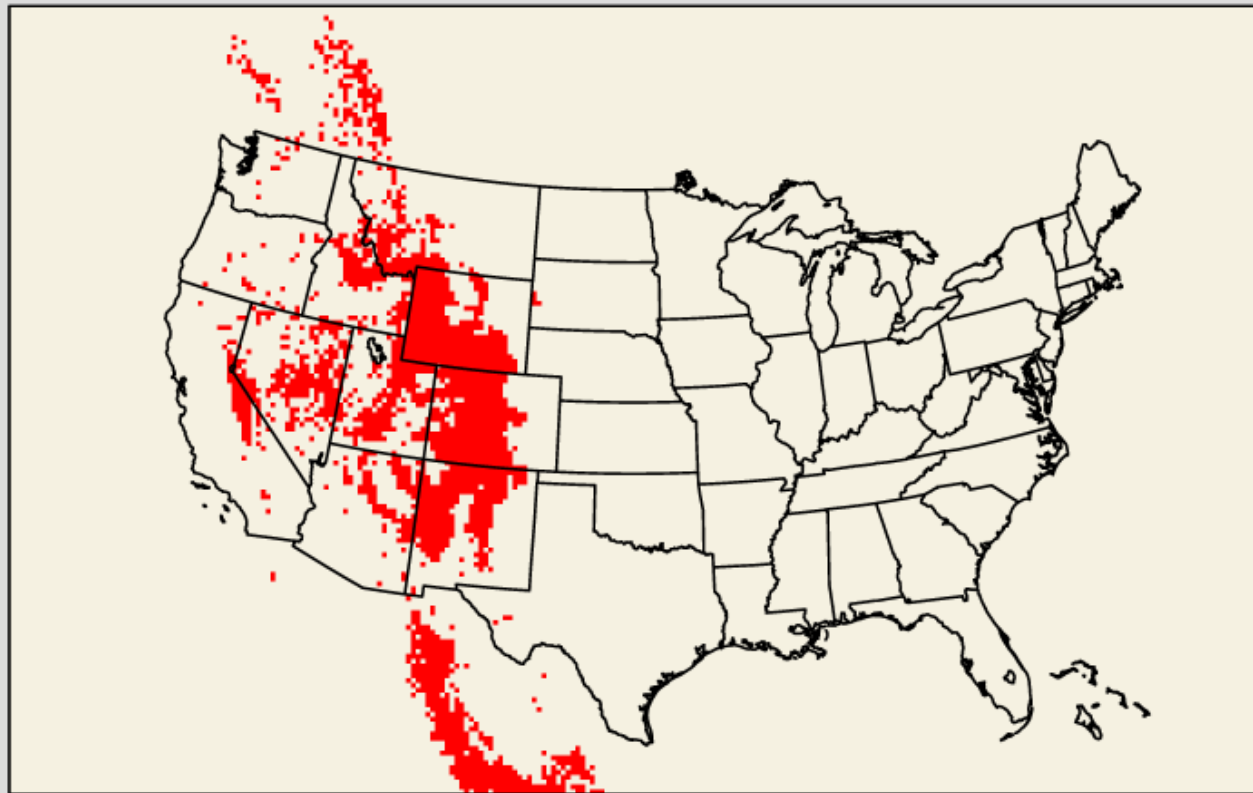


NCEP grids link: www.nco.ncep.noaa.gov/pmb/docs/on388/tableb.html

Station Masking: The user can specify a list of stations and Point-Stat will verify only at those stations.



Data Threshold Masking: The user may specify a threshold on some other data field for masking.



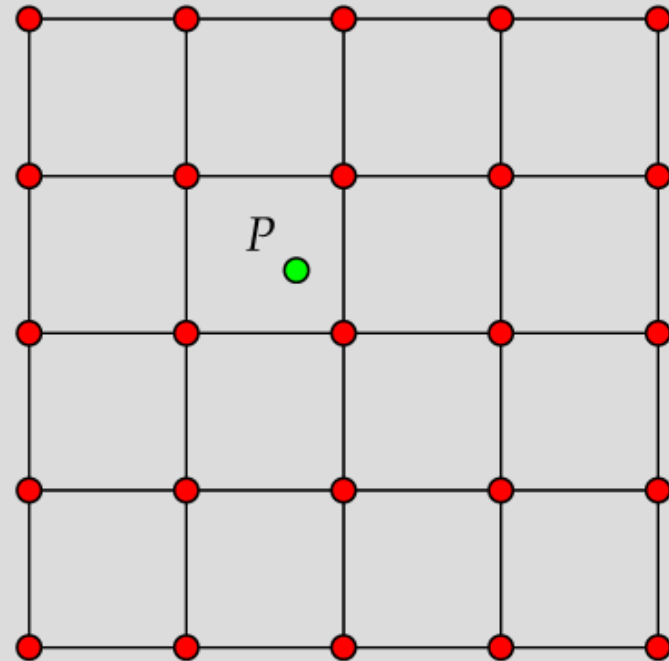
Topography > 6000 feet

Interpolation

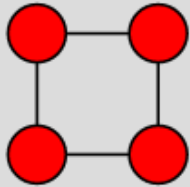
Need to Choose:

(1) Method

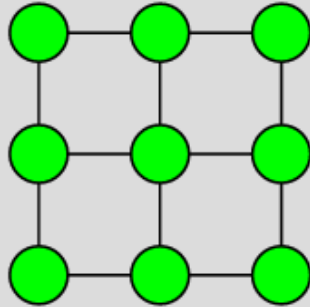
(2) Width



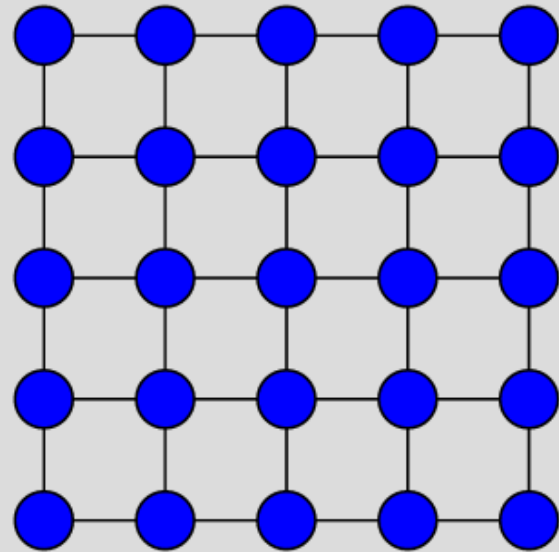
Interpolation Widths



$N = 2$



$N = 3$



$N = 5$

Interpolation Methods

	Min	Max	Median	UW Mean	DW Mean	Nearest Nbr	Least Squares
Point Stat	✓	✓	✓	✓	✓	✓	✓
Grid Stat	✓	✓	✓	✓	N/A	N/A	N/A

For Grid Stat, these are smoothing methods.

Min, Max, Median

Takes minimum, maximum or median of values in interpolation square.

Median separates the upper half of data values from the lower half. This is different from the mean, which is an average.

Nearest Neighbor

Essentially, no interpolation is performed.

Value at interpolation point is simply the data value at the closest grid point.

Unweighted Mean Distance-Weighted Mean

Unweighted Mean is the average.

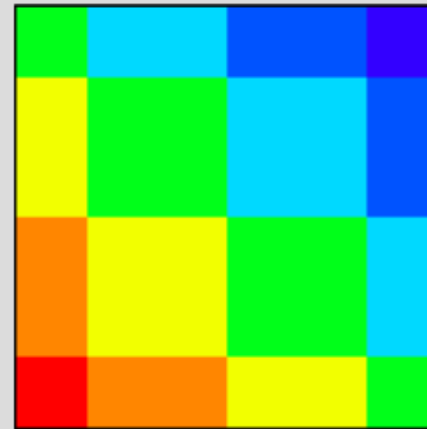
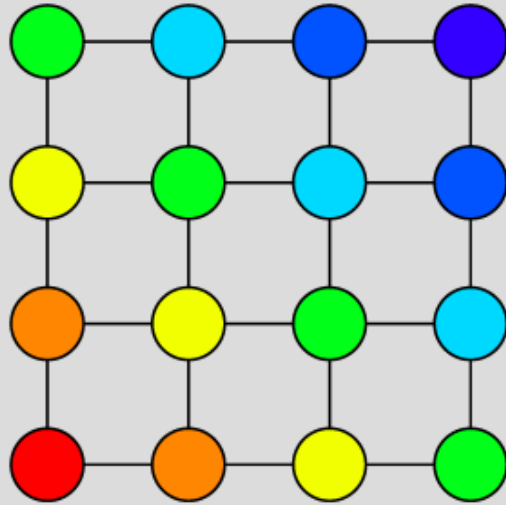
Distance-Weighted Mean is an average weighted according to distance from nearby grid points.

Least Squares

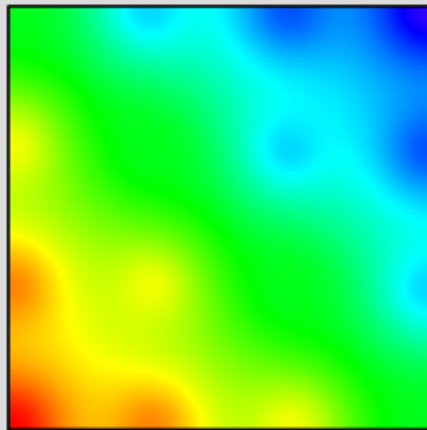
Performs a local Least-Squares linear fit in interpolation square.

$$z = Ax + By + C$$

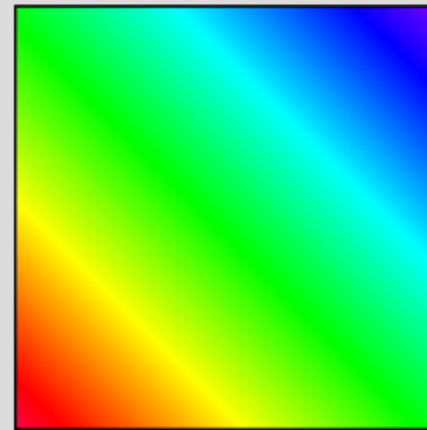
Interpolation Examples



Nearest Neighbor



Distance Weighted Mean



Least Squares

Interpolation Shape

- Add support for circular interpolation, regridding, and data smoothing areas rather than just squares.
- Smooth data by computing the **MAX** of **WIDTH** of 30.

