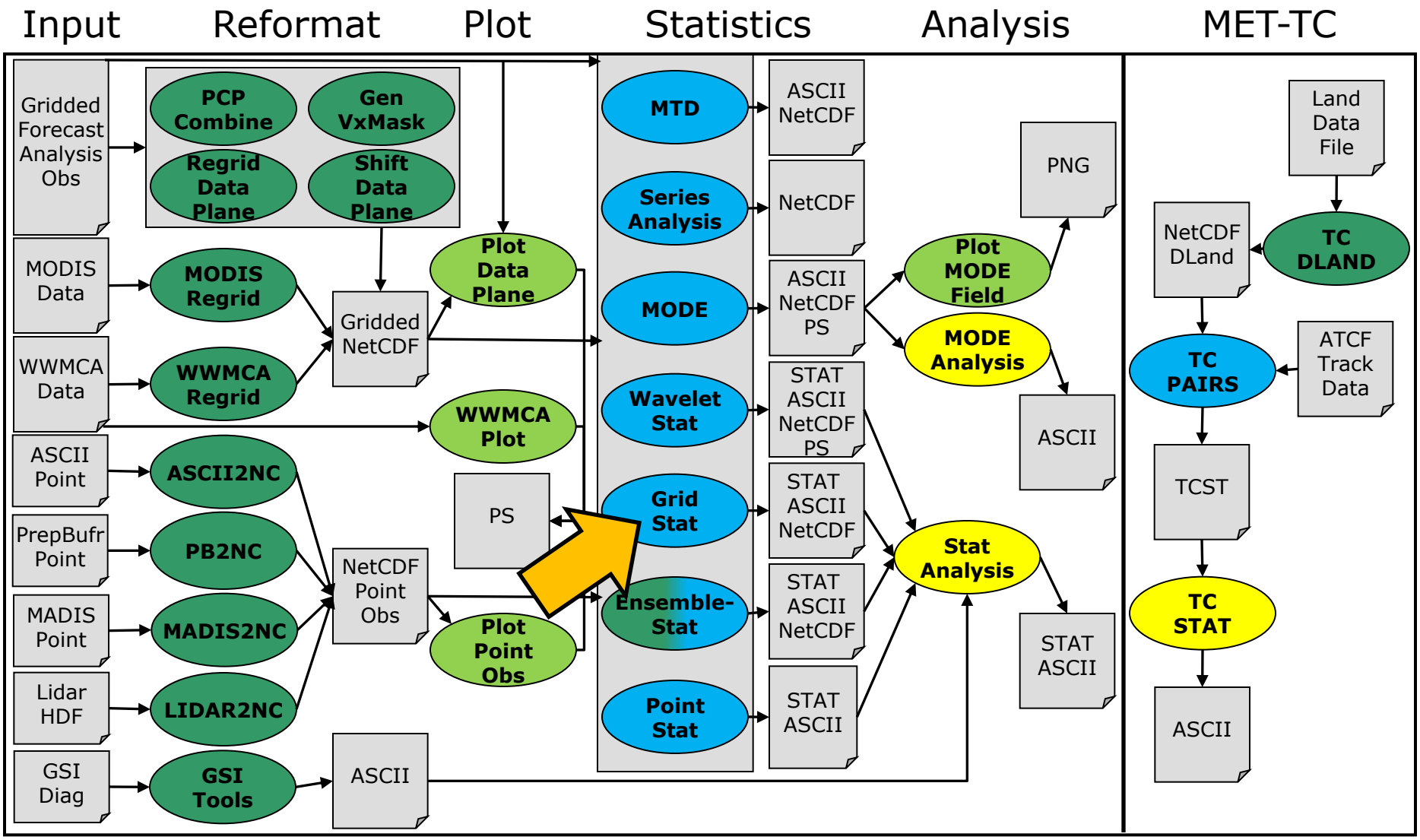
MET Grid-Stat Tool

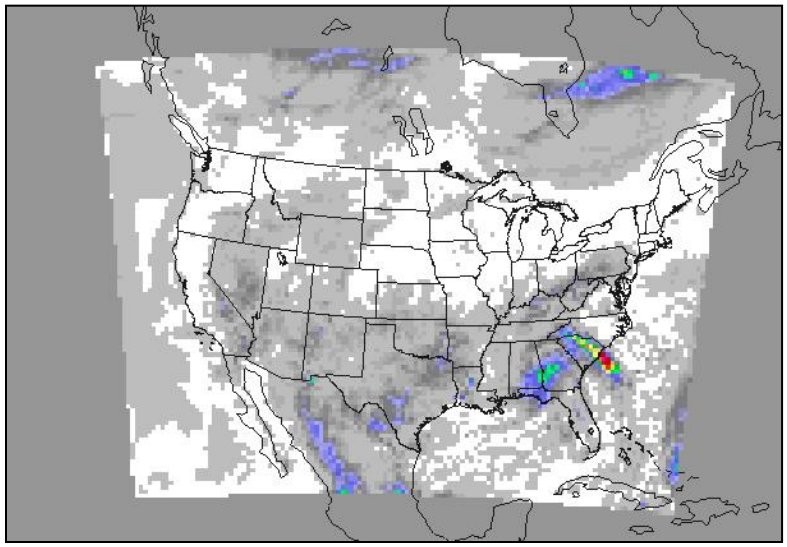
John Halley Gotway

METplus Tutorial
July 31 – August 2, 2019
NRL – Monterey, CA

Grid-Stat Tool

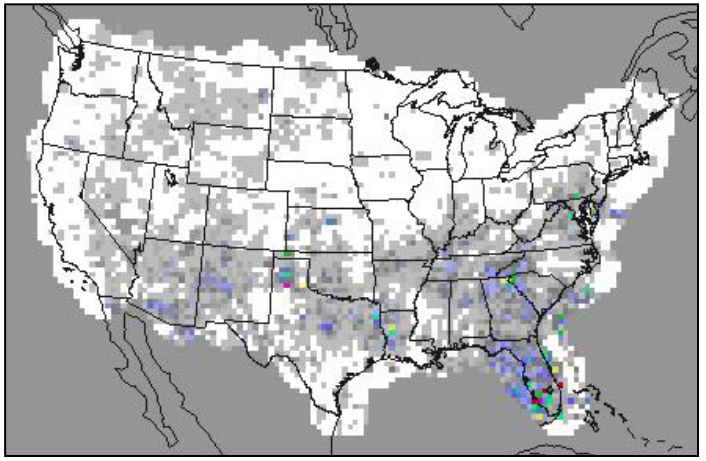


Grid-Stat: Overview



- Compare **gridded forecasts** to **gridded observations** on the same grid.
- 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.

- 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.
- Methods for probabilistic forecasts.
- Economic Cost/Loss Value.
- Neighborhood verification methods.
- Fourier decomposition.
- Gradient statistics.



Grid-Stat: Usage

Usage: grid_stat

fcst_file

obs_file

config_file

[-outdir path]

[-log file]

[-v level]

[-compress level]

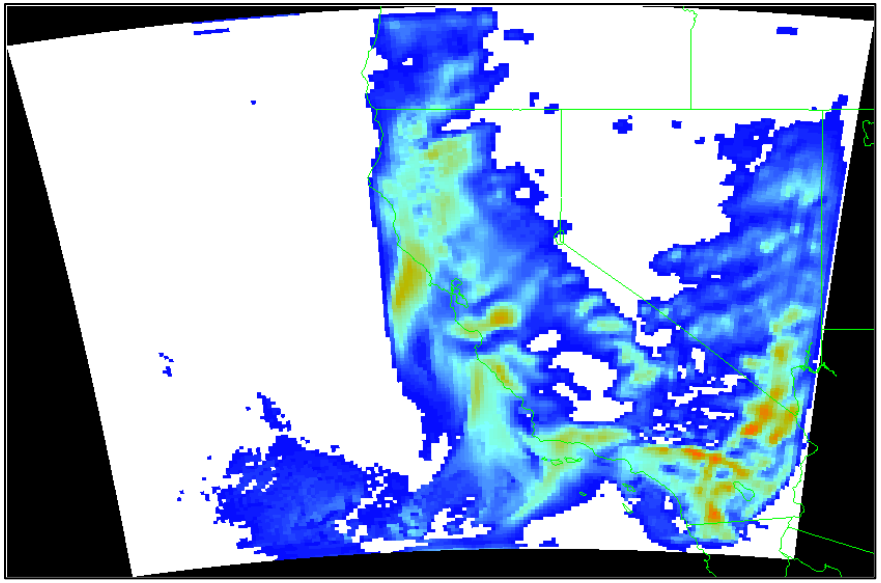
fcst_file	Gridded forecast file
obs_file	Gridded observation file
config_file	ASCII configuration file
-outdir	Output directory to be used
-log	Optional log file
-v	Level of logging

Grid-Stat: Input/Output

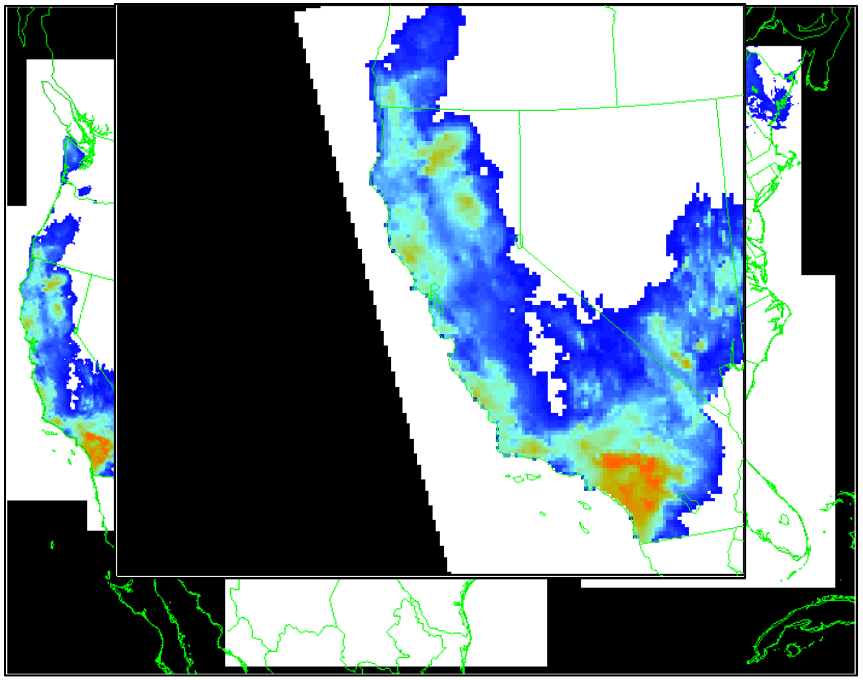
- Input Files
  - Gridded forecast and observation files
    - GRIB1 output of Unified Post-Processor (or other)
    - GRIB2 from NCEP (or other)
    - NetCDF from PCP-Combine, wrf_interp, or CF-compliant
    - Python Interface
  - ASCII configuration file
- Output Files
  - ASCII statistics file with all output lines (ends with ".stat")
  - Optional ASCII files sorted by line type with a header row (ends with "_TYPE.txt")
  - Optional NetCDF matched pairs file (ends with "_pairs.nc")

Grid-Stat: Common Grid

Model Forecast



StageIV Analysis

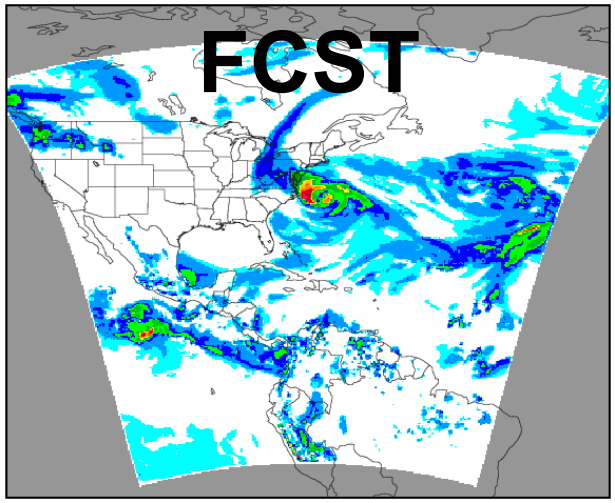


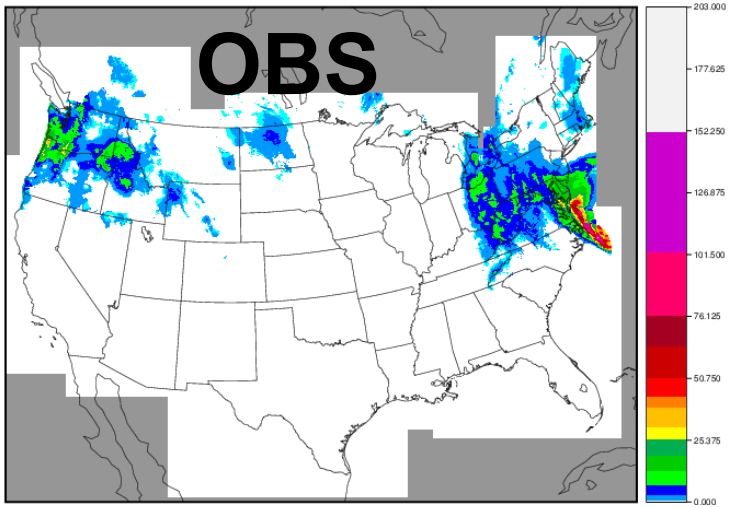
- Forecast and observations must be placed on a common grid.
- Regrid the StageIV Analysis (GRIB) to the model domain:

```
copygb –xg"255 5 169 154 31357 -129770 8 -120500 10395 10395 0 64" \
   ST4.2010122212.06h ST4.2010122212.06h_regrid
```

- Automated regridding in **configuration file** or use **regrid_data_plane**.

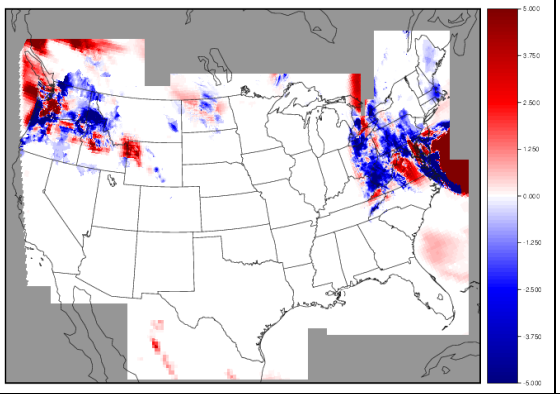
Grid-Stat: Automated Regriding



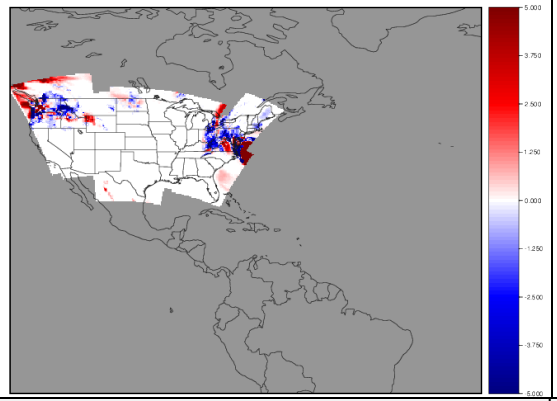



```
//
// Verification grid
//
regrid = {
   to_grid     = NONE;
   method      = BUDGET;
   width       = 2;
   vld_thresh  = 0.5;
   shape       = SQUARE;
}
```

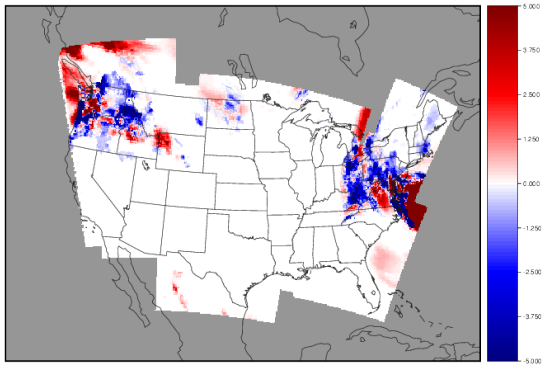
DIFF: FCST - OBS



`to_grid = FCST;`



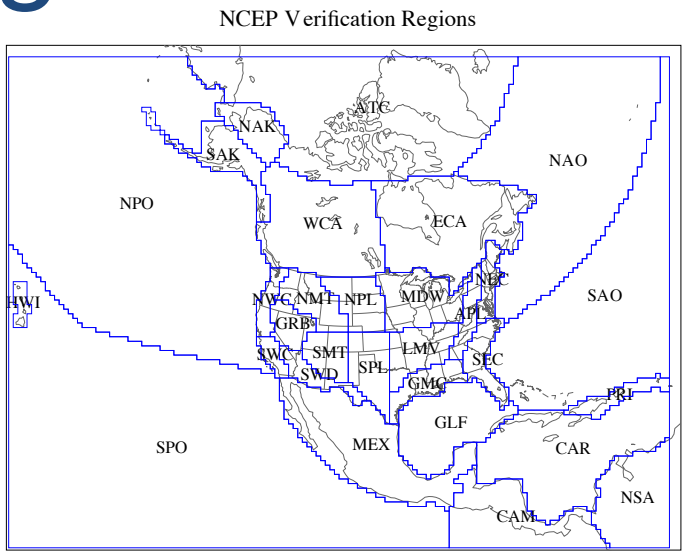
`to_grid = OBS;`

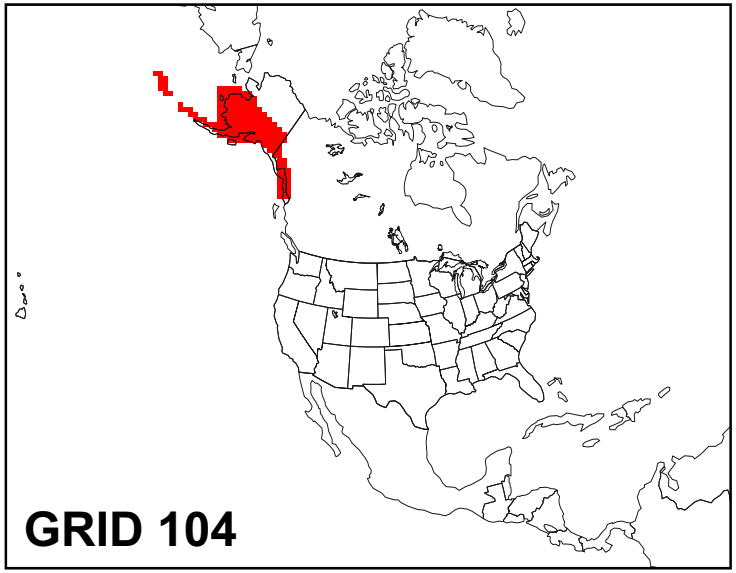


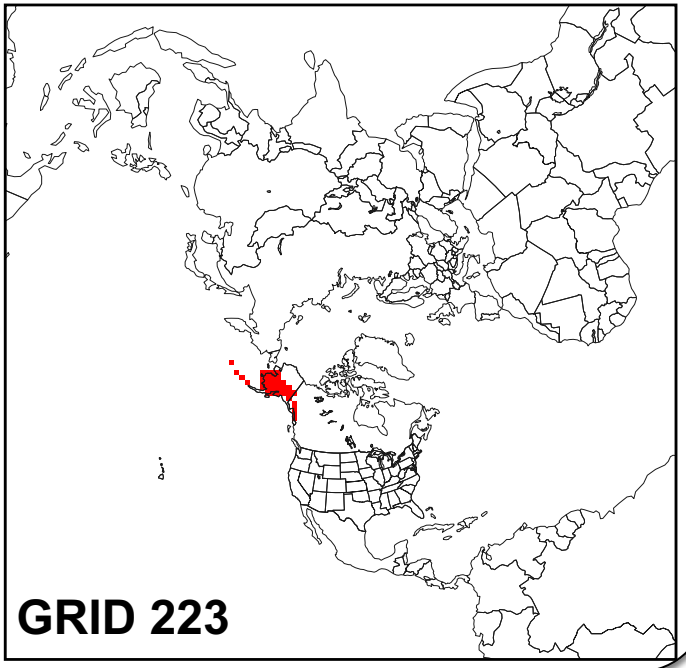
`to_grid = "G130";`

Automated Regridding of Masks

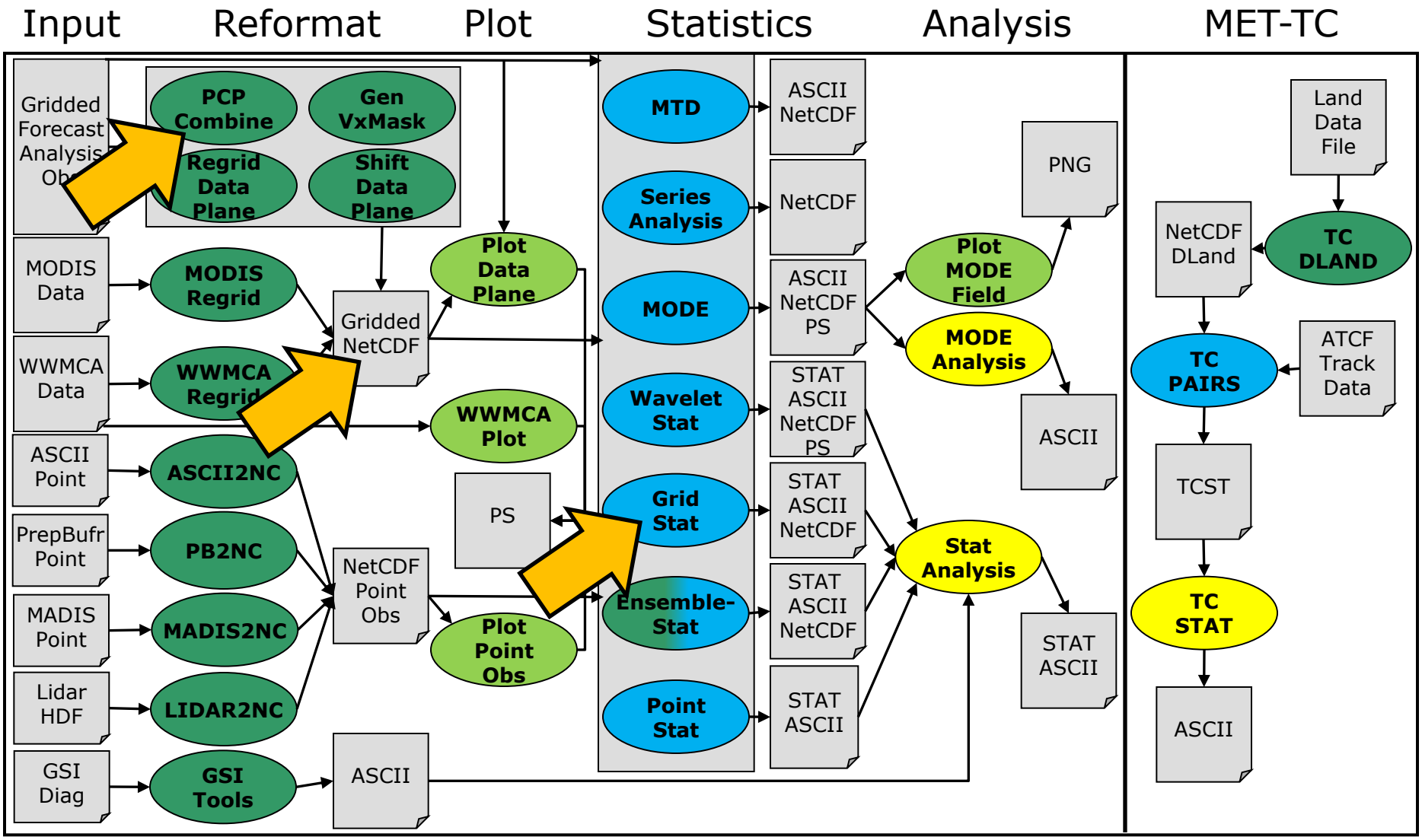
- NCEP Verification Regions defined for Grid 104.
  - CONUS, EAST, and WEST are groups of sub-regions.
- MET release includes polyline files (data/poly) and **NetCDF masks** (data/poly/NCEP_masks).
- Automatically regrid mask (NEAREST neighbor) to verification domain:
  - *DEBUG 2: Regridding mask grid to the verification grid using nearest neighbor interpolation…*







Example Data Flow



Grid-Stat: Configuration

- Many configurable parameters – only set a few:
  - Precipitation accumulated over 24 hours.
    - GRIB1 forecast
    - NetCDF observation
  - Threshold any rain and moderate rain (mm).
  - Accumulate stats over all the points in the domain and just the eastern United States.
  - Compute neighborhood statistics with two sizes.
  - Generate continuous, categorical, and neighborhood line types.
  - Write NetCDF pairs.

```
fcst = { // GRIB Input
 field = [
  {
   name       = "APCP";
   level      = [ "A24" ];
   cat_thresh = [ >0.0, >20.0 ];
  }
 ];
}
```

```
obs = { // NetCDF from PCP-Combine
 field = [
  {
   name       = "APCP_24";
   level      = [ "(*,*)" ];
   cat_thresh = [ >0.0, >20.0 ];
  }
 ];
}
```

```
mask = {
 grid = [ "FULL" ];
 poly = [
  "MET_BASE/poly/EAST.poly" ];
}
```

```
nbrhd = {
 vld_thresh = 1.0;
 width      = [ 3, 5 ];
 cov_thresh = [ >=0.5 ];
}
```

```
output_flag = {
 fho     = BOTH;
 ctc     = BOTH;
 cts     = BOTH;
 mctc    = BOTH;
 mcts    = BOTH;
 cnt     = BOTH;
 sl1l2   = BOTH;
 sal1l2  = NONE;
 vl1l2   = NONE;
 val1l2  = NONE;
 pct     = NONE;
 pstd    = NONE;
 pjc     = NONE;
 prc     = NONE;
 eclv    = NONE;
 nbrctc  = BOTH;
 nbrcts  = BOTH;
 nbrcnt  = BOTH;
 grad    = NONE;
}
```

```
nc_pairs_flag = {
   latlon      = TRUE;
   raw         = TRUE;
   diff        = TRUE;
   climo       = TRUE;
   weight      = FALSE;
   nbrhd       = FALSE;
   fourier     = FALSE;
   gradient    = FALSE;
   apply_mask  = TRUE;
}
```

Grid-Stat: Field Name and Level

- GRIB1 and GRIB2 files
  - **name = "GRIB Abbreviation";**
    - http://www.nco.ncep.noaa.gov/pmb/docs/on388/table2.html
    - TMP for Temperature, APCP for accumulated precipitation.
  - **level = ["string"];** Multiple values expand to multiple vx tasks
    - Level indicator followed by level value.
      - **A** for accumulation interval in HH[MMSS] format *(A06)*.
      - **P** for pressure level (*P500*) or layer (*P500-600*).
      - **Z** for vertical level (*Z2* or *Z10*).
      - **L** for generic level type (*L100*).
      - **R** for a specific GRIB record number (*R225*).
- Gridded NetCDF files
  - **name = "string";** Defines NetCDF variable name.
  - **level = ["string"];** Defines index into dimensions.
  - For APCP_06(lat,lon) from PCP-Combine output
    - *name = "APCP_06"; level = ["(*,*)"];*
  - For TT(Time, num_metgrid_levels, south_north, west_east) from p_interp
    - *name = "TT"; level = ["(0,0,*,*)", "(0,1,*,*)", "(0,2,*,*)"];*

Grid-Stat: Config File Defaults

- MET Statistics tools parse up to 4 configuration files:
  1. **MET_BASE/config/ConfigConstants** defines configuration file constants (e.g. NONE, STAT, BOTH) and should not be modified.
  2. **MET_BASE/config/ConfigMapData** defines default map data for all plots (map data files, line colors, widths, and types for Plot-Point-Obs, Plot-Data-Plane, Wavelet-Stat, and MODE).
  3. **MET_BASE/config/GridStatConfig_default** defines default settings for the specific tool.
  4. **User-specific configuration** file passed on the command line override default settings.

NOTE: **MET_BASE/config/README** describes config file options.

NOTE: When running a shared installation of MET, override default settings in the **user-specific configuration** file rather than modifying the system-wide defaults.

Grid-Stat: Run

- grid_stat \
  sample_fcst.grb sample_obs.nc \
  GridStatConfig_APCP24 -outdir out -v 2

```
DEBUG 1: Default Config File: met-X.Y/share/met/data/config/GridStatConfig_default
DEBUG 1: User Config File: GridStatConfig_APCP24
DEBUG 1: Forecast File: sample_fcst.grb
DEBUG 1: Observation File: sample_obs.nc
DEBUG 2: ----------------------------------------------------------------------
DEBUG 2: Processing APCP/A24 versus APCP_A24, for interpolation method UW_MEAN(1), over region FULL, using 6412 pairs
DEBUG 2: Computing Categorical Statistics.
DEBUG 2: Computing Multi-Category Statistics.
DEBUG 2: Computing Continuous Statistics.
DEBUG 2: Processing APCP/A24 versus APCPA24, for interpolation method UW_MEAN(1), over region EAST, using 2582 pairs.
DEBUG 2: Processing APCP/A24 versus APCPA24, for interpolation method NBRHD(9), raw thresholds of >0.000 and >0.000,
over region EAST, using 5829 pairs.
DEBUG 2: Computing Neighborhood Categorical Statistics.
DEBUG 2: Computing Neighborhood Continuous Statistics.
… MORE NEIGHBORHOOD VERIFICATION TASKS LISTED …
DEBUG 2: ----------------------------------------------------------------------
DEBUG 1: Output file: out/grid_stat_240000L_20050808_000000V.stat
DEBUG 1: Output file: out/grid_stat_240000L_20050808_000000V_fho.txt
DEBUG 1: Output file: out/grid_stat_240000L_20050808_000000V_ctc.txt
DEBUG 1: Output file: out/grid_stat_240000L_20050808_000000V_cts.txt
DEBUG 1: Output file: out/grid_stat_240000L_20050808_000000V_mctc.txt
DEBUG 1: Output file: out/grid_stat_240000L_20050808_000000V_mcts.txt
DEBUG 1: Output file: out/grid_stat_240000L_20050808_000000V_cnt.txt
DEBUG 1: Output file: out/grid_stat_240000L_20050808_000000V_sl1l2.txt
DEBUG 1: Output file: out/grid_stat_240000L_20050808_000000V_nbrctc.txt
DEBUG 1: Output file: out/grid_stat_240000L_20050808_000000V_nbrcts.txt
DEBUG 1: Output file: out/grid_stat_240000L_20050808_000000V_nbrcnt.txt
DEBUG 1: Output file: out/grid_stat_240000L_20050808_000000V_pairs.nc
```

Grid-Stat: ASCII Output

- 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)
- Wind 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)
- Neighborhood – apply threshold, define neighborhood
  - Neighborhood continuous statistics (NBRCNT)
  - Neighborhood contingency table counts (NBRCTC)
  - Neighborhood contingency table statistics (NBRCTS)
- Gradient line type (GRAD)

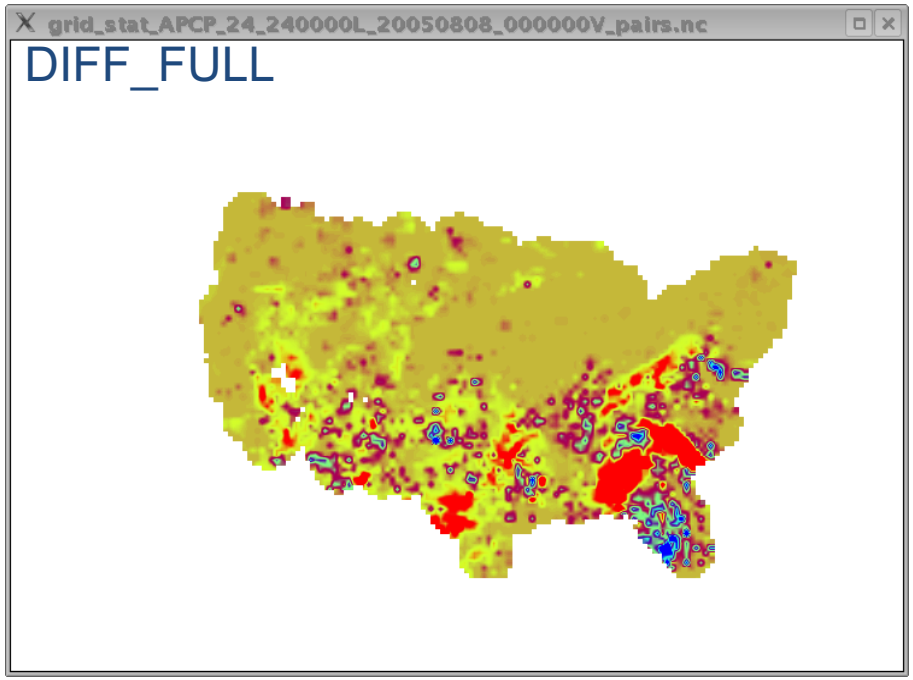
24 common header columns + Line type specific columns

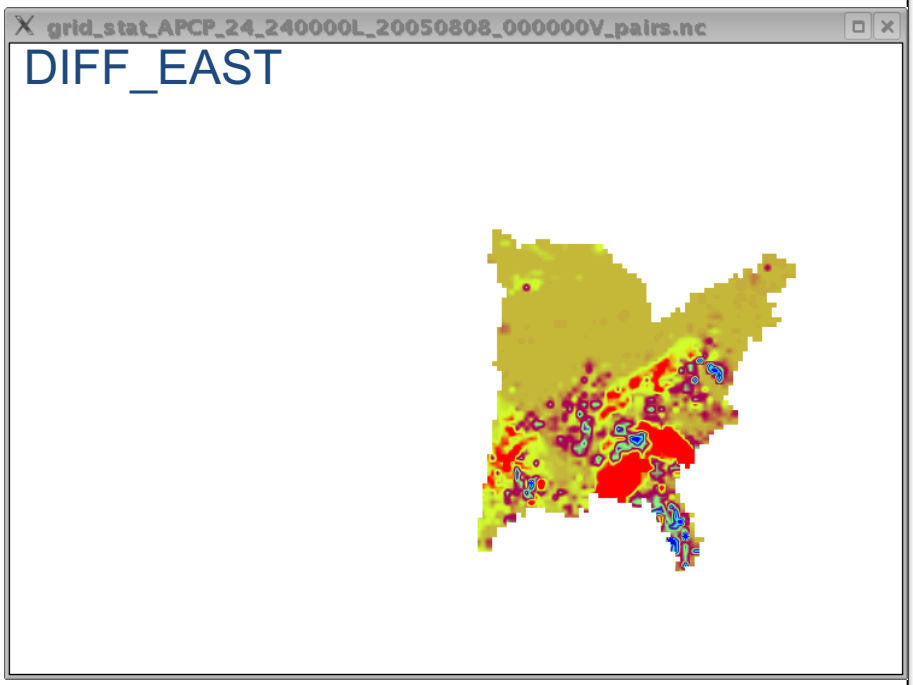
Grid-Stat: Sample Output

1. STAT file output for sample run:
   - 2 lines each for CNT, MCTC, MCTS, and SL1L2
     - = 2 verification regions (FULL and EAST)
   - 4 lines each for FHO, CTC, and CTS
     - = 2 regions * 2 thresholds
   - 8 lines each for NBRCNT, NBRCTC, NBRCTS
     - = 2 regions * 2 thresholds * 2 neighborhood sizes
2. Additional TXT files for each line type
3. NetCDF file containing matched pairs

Grid-Stat: NetCDF Matched Pairs

- Forecast, observation, and difference fields for each combination of…
  - Variable, level, masking region, and interpolation method (smoothing)
- Sample output contains 6 fields:
  - FCST, OBS, and DIFF for FULL and EAST





- Set **`apply_mask = FALSE;`** for only FULL domain.

Grid-Stat: CTC Output Line

VERSION	VX.Y
MODEL	WRF
DESC	NA
FCST_LEAD	240000
FCST_VALID_BEG	20050808_000000
FCST_VALID_END	20050808_000000
OBS_LEAD	000000
OBS_VALID_BEG	20050808_000000
OBS_VALID_END	20050808_000000
FCST_VAR	APCP_24
FCST_UNITS	mm
FCST_LEV	A24
OBS_VAR	APCP_24
OBS_UNITS	mm
OBS_LEV	A24

OBTYPE	MC_PCP
VX_MASK	EAST
INTERP_MTHD	UW_MEAN
INTERP_PNTS	1
FCST_THRESH	>20.000
OBS_THRESH	>20.000
COV_THRESH	NA
ALPHA	NA
LINE_TYPE	CTC
TOTAL	2582
FY_OY (hits)	5
FY_ON (f.a.)	104
FN_OY (miss)	70
FN_ON (c.n.)	2403

FHO vs CTC Line Type

- Grid-Stat, Point-Stat, and Stat-Analysis can output FHO and CTC line types.
  - Values are equivalent (*and redundant*).
  - CTC has integer counts for 4 cells of 2x2 table.
  - FHO has floating point rates.
  - FHO rounding issues for rare events.

LINE_TYPE	CTC
TOTAL	2582
FY_OY (hits)	5
FY_ON (false alarms)	104
FN_OY (misses)	70
FN_ON (correct neg)	2403

LINE_TYPE		FHO
TOTAL		2582
F_RATE (fcst rate)	(5+104)/2582	0.042215
H_RATE (hit rate)	5/2582	0.0019365
O_RATE (obs rate)	(5+70)/2582	0.029047

Comparing Different Fields

- Grid-Stat, Point-Stat, and all STAT tools may be used to compare different variables.
  - User must interpret results.
  - Example: Convective Precip vs. Total Precip
  - Configuration file settings:
    - Selecting variable/levels

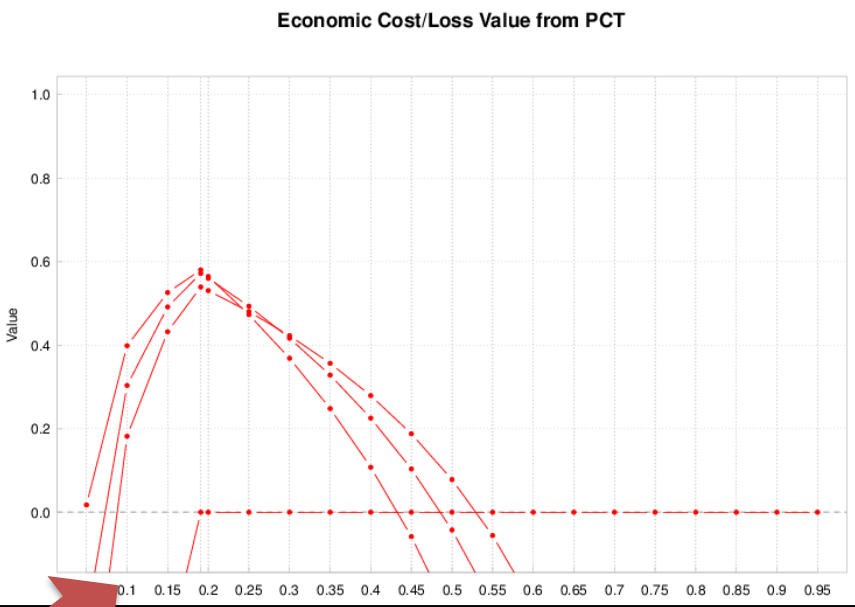
```
fcst = {
 field = [
  {
   name        = "ACPCP";
   level       = [ "A24" ];
   cat_thresh = [ >0.0 ];
  }
 ];
};
```

```
obs = {
 field = [
  {
   name        = "APCP";
   level       = [ "A24" ];
   cat_thresh = [ >0.0 ];
  }
 ];
};
```

Economic Cost/Loss Value

- Grid-Stat, Point-Stat, and Stat-Analysis can output the ECLV line type.
- Equivalent to the VSDB ECON line type, except…
  - ECON is only generated when evaluating ensemble probabilities.
  - ECLV from 2x2 CTC contingency table yields a single curve.
  - ECLV from Nx2 PCT probabilistic contingency table yields N curves.
- One ECLV line equals one curve on the plot.
  - Undefined at 0 and 1.
  - Maximized for the base rate.

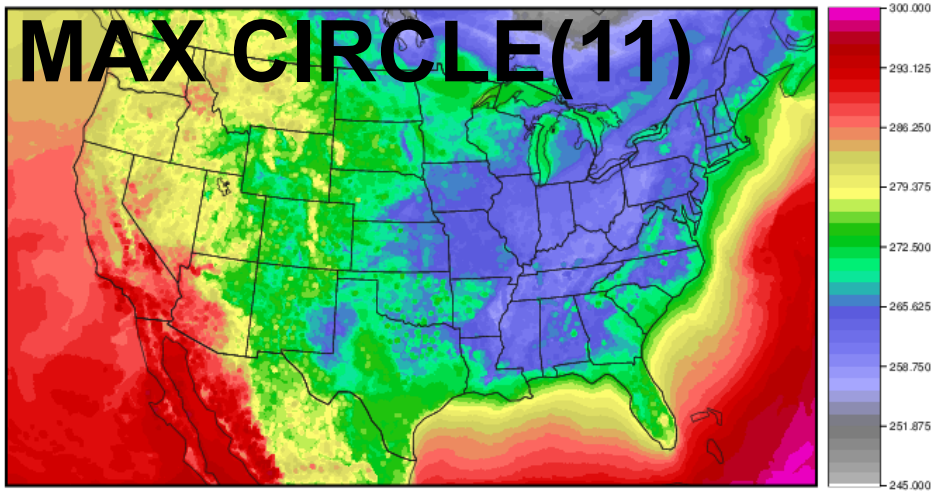
Forecast	Observation		Total
	o = 1 (e.g., "Yes")	o = 0 (e.g., "No")	
p_1 = midpoint of (0 and threshold1)	n_{11}	n_{10}	$n_{1\cdot} = n_{11} + n_{10}$
p_2 = midpoint of (threshold1 and threshold2)	n_{21}	n_{20}	$n_{2\cdot} = n_{21} + n_{20}$
⋮	⋮	⋮	⋮
p_i = midpoint of (threshold$i$ and 1)	n_{i1}	n_{i0}	$n_{i\cdot} = n_{j1} + n_{j0}$
Total	$n_{\cdot 1} = \Sigma n_{i1}$	$n_{\cdot 0} = \Sigma n_{i0}$	$T = \Sigma n_i$




```
eclv_points = 0.05; // 0 to 1 every 0.05
eclv_points = [ 0.1, 0.2, 0.3, 0.4, 0.5, 0.65, 0.8, 0.95 ]; // non-equal
```

Grid-Stat: Data Smoothing

- Higher resolution forecasts typically score worse than lower resolution ones for traditional scores, like RMSE.
- Specify **interp** section to apply smoothing method(s) prior to computing statistics.
- Smoothing methods indicated in **INTERP_MTHD** and **INTERP_PNTS** columns.

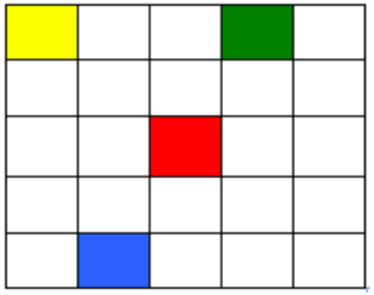


hrrr_TMP_Z2_max_121.nc

```
//
// Data smoothing methods
//
interp = {
  // FCST, OBS, or BOTH
  field      = BOTH;
  vld_thresh = 1.0;
  shape      = SQUARE;
  type = [
    // Default, no smoothing
    { method = NEAREST;
      width  = 1; },
    // Mean of 11x11 square
    { method = UW_MEAN;
      width  = 11; },
    // Max of circle diam 11
    { method = MAX;
      width  = 11;
      shape  = CIRCLE; }
  ];
}
```

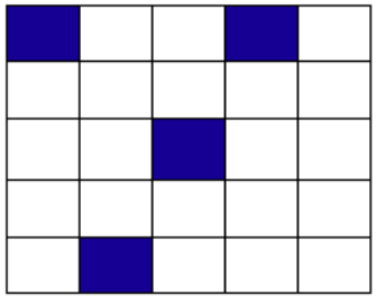
Grid-Stat: Neighborhoods

- As with all neighborhood methods (i.e. HiRA), allows for some spatial / temporal uncertainty in either model or observation by giving credit for being 'close'.
- Apply categorical threshold and neighborhood width to convert gridded forecast and observation fields into fractional coverage fields.
- Select SQUARE or CIRCLE shape.
- Every permutation of **cat_thresh** and **nbrhd.width**.
  - **NBRCNT** statistics (**FBS**, **FSS**) computed directly from fractional coverage fields.
  - Apply **cov_thresh** thresholds to fractional coverage fields to compute **NBRCTC** and **NBRCTS** counts and statistics (i.e. like **CTC** and **CTS**).



Model Forecast
White boxes = 0
Colored boxes > 0

Threshold Forecast
Blue boxes = event



Fractional Proportion
- 1x1 Neighborhood: 1/1
- 3x3 Neighborhood: 1/9
- 5x5 Neighborhood: 4/25

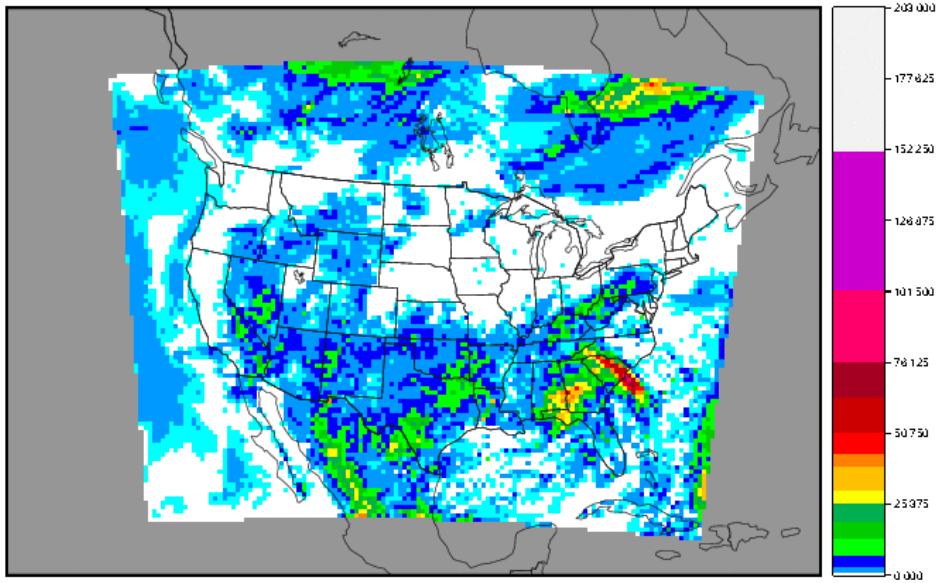
```
cat_thresh = [ >0.0, >=6.35 ];
...
nbrhd = {
   width      = [ 1, 3, 5, 7 ];
   cov_thresh = [ >=0.5 ];
   vld_thresh = 1.0;
   shape      = SQUARE;
}
```

Grid-Stat: Neighborhoods

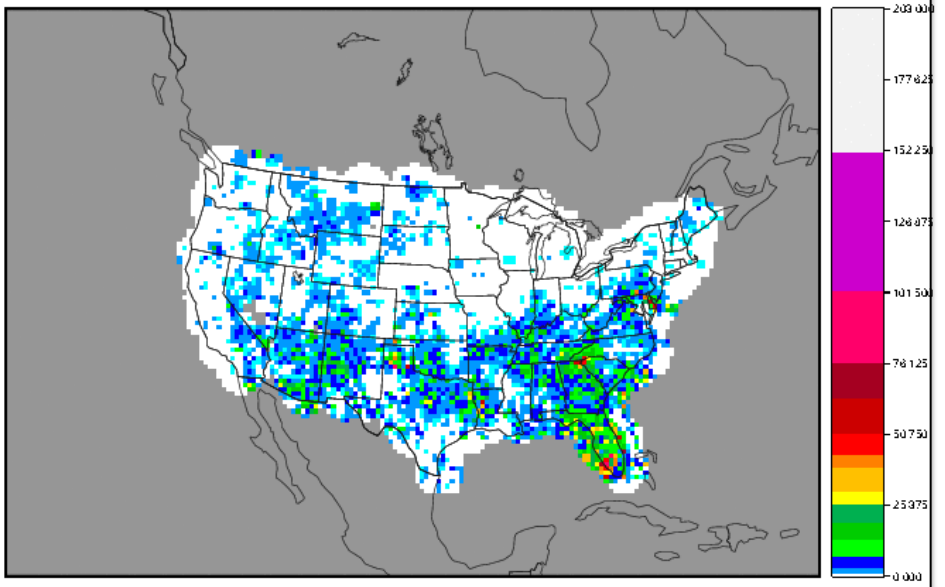
- Threshold 24-APCP >= 0.1”
- Edge effects as width increases.
- FSS increases as width increases:
  - 0.64287, 0.73593, 0.80247
  - 0.85106, 0.89191, 0.91487
  - 0.92632, 0.93536, 0.94517

```
cat_thresh = [ >=2.54 ];

nbrhd = {
   width = [ 3,  5,  7,  9, 11,
             13, 15, 17, 19 ];
}
nc_pairs_flag = {
   nbrhd = TRUE;
}
```



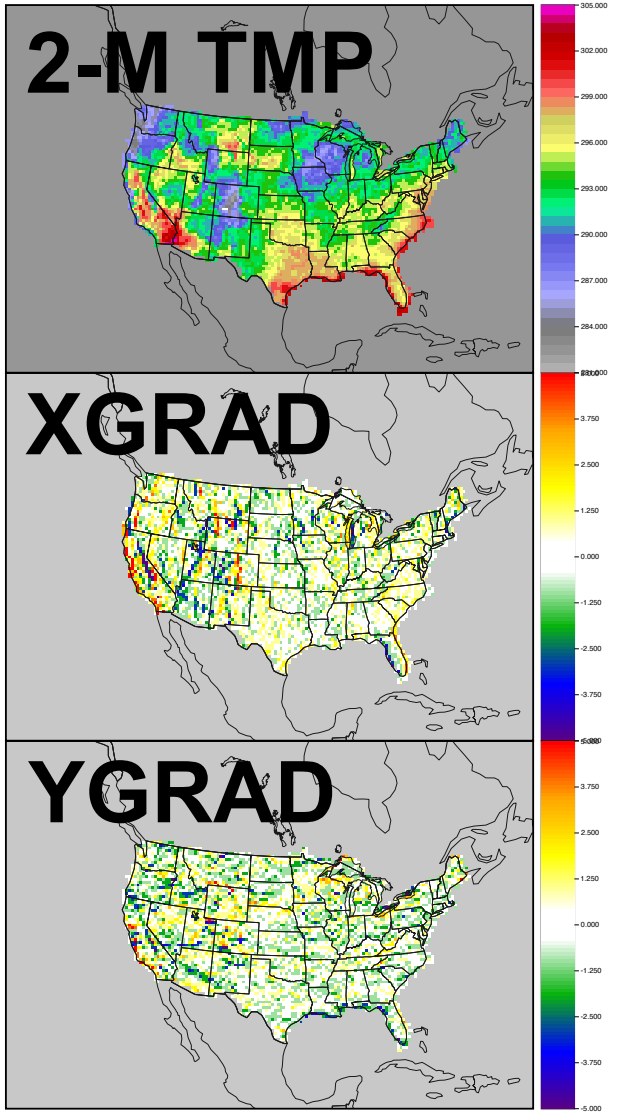
grid_stat_APCP_24_240000L_20050808_000000V_pairs.nc



grid_stat_APCP_24_240000L_20050808_000000V_pairs.nc

Grid-Stat: Gradients

- **GRAD** line type contains the S1 score and its components.
  - WMO-mandated statistic from 1954.
  - Computed over the gradients of forecast and observation fields computed in the X and Y grid direction.
  - Adapted from VSDB code:
    - FGBAR: mean forecast gradient
    - OGBAR: observed gradient
    - MGBAR: mean of maximum gradient
    - EGBAR: mean of gradient differences
    - S1 = 100 * EGBAR / MGBAR
    - S1_OG = 100 * EGBAR / OGBAR
    - FGOG_RATIO = FGBAR / OGBAR

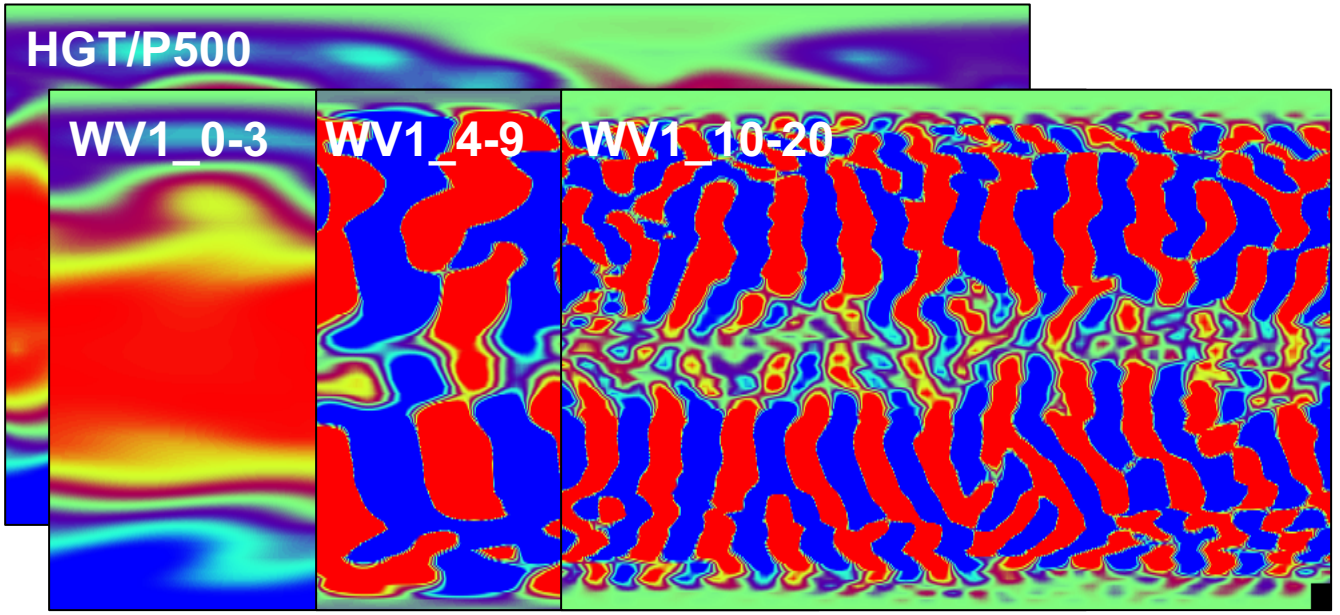


Grid-Stat: Fourier Decomposition

- Supports 1-Dimensional Fourier decompositions.
- Affects output for **CNT, SL1L2, SAL1L2, VL1L2**, and **VAL1L2** line types.
- Configuration file option to specify the waves:

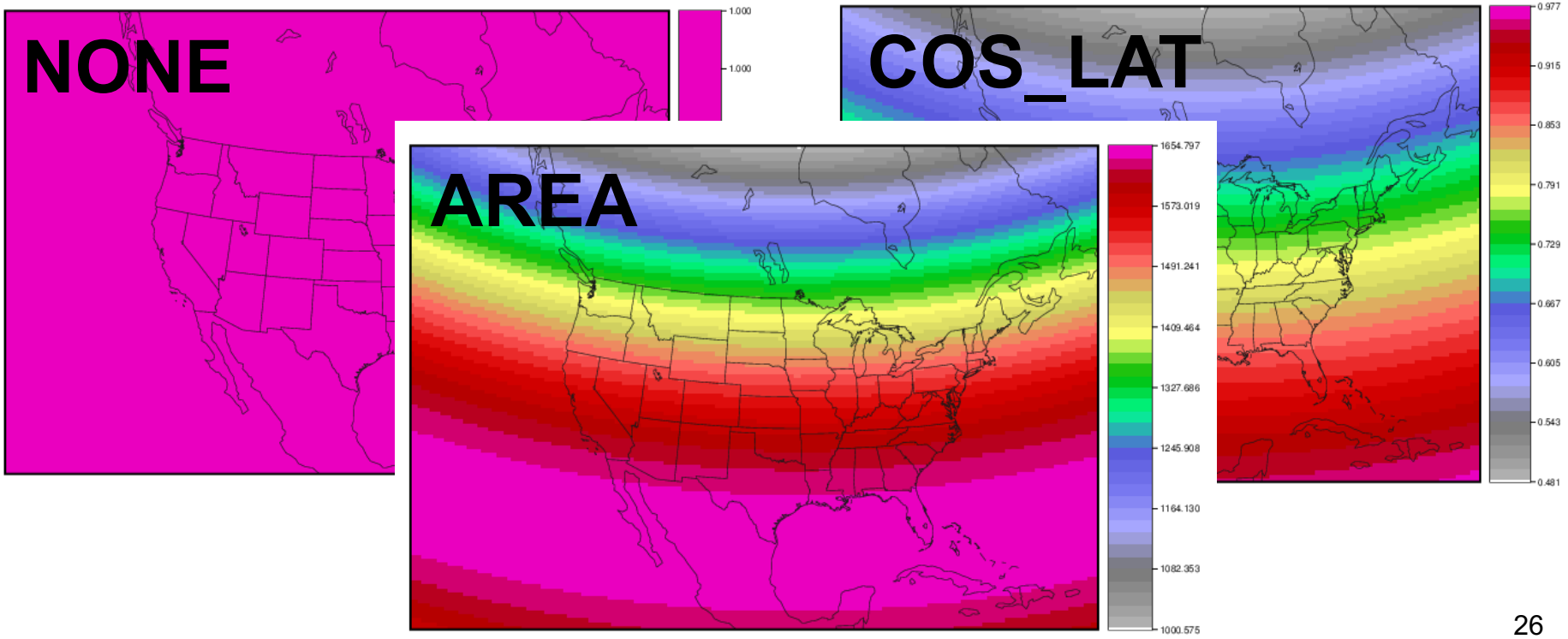
```
fourier = {
  wave_1d_beg = [  0, 0, 4, 10 ];
  wave_1d_end = [ 72, 3, 9, 20 ];
}
```

- Wave numbers indicated in the **INTERP_MTHD** column:
  - **WV1_0-72**
  - **WV1_0-3**
  - **WV1_4-9**
  - **WV1_10-20**



Grid-Stat: Grid Box Weighting

```
//
// The "grid_weight_flag" specifies how grid weighting should be applied…
//    - "NONE" to disable grid weighting using a constant weight (default).
//    - "COS_LAT" to define the weight as the cosine of the grid point latitude.
//      This an approximation for grid box area used by NCEP and WMO.
//    - "AREA" to define the weight as the true area of the grid box (km^2).
//
grid_weight_flag = NONE;
```



Config: Conditional Continuous

- Continuous statistics (CNT) from Grid-Stat and Point-Stat include FBAR, OBAR, ME, MAE, RMSE, and PR_CORR.
- A single matched pair (MPR) consists of an observation value paired with a forecast value.
- By default, all pairs within the masking region are included in the CNT statistics.
- Specify **cnt_thresh** to filter which pairs should be included.
- Specify **cnt_logic** to combine filters for fcst and obs data.
- These settings are listed in the FCST_THRESH and OBS_THRESH columns.

```
//
// fcst OR obs meets threshold
//
cnt_thresh = [ NA, >1.0, >3.0 ];
cnt_logic  = UNION;
fcst = { … }
obs  = { … } * NA threshold always TRUE
```

```
//
// fcst AND obs meets threshold
//
cnt_thresh = [ NA, >1.0, >3.0 ];
cnt_logic  = INTERSECTION;
fcst = { … }
obs  = { … }
```

```
//
// obs meets threshold
//
cnt_logic = INTERSECTION;
fcst = { cnt_thresh = [ NA, NA, NA ]; }
obs  = { cnt_thresh = [ >1.0, >3.0 ]; }
```

Config: Converting Data

- Config file language supports functions of 1 variable.
- Use convert(x) function to define unit conversions:

```
convert(x) = log10(x);
convert(x) = sqrt(x);
```

- Common conversion functions pre-defined in **share/met/config/ConfigConstants**:

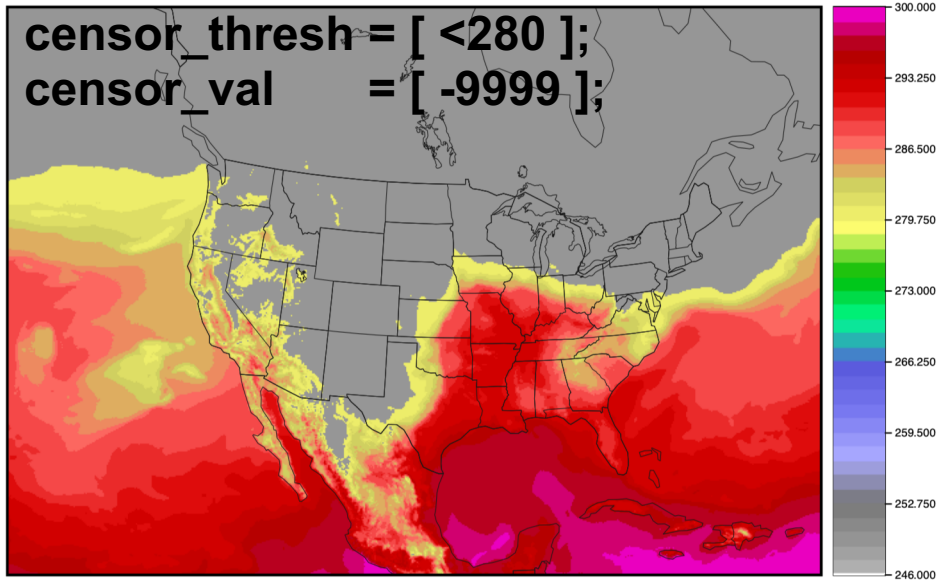
```
K_to_C(t) = t - 273.15;
C_to_K(t) = t + 273.15;
C_to_F(t) = 1.8 * t + 32.0;
```

Config: Censoring Data

- Applying MET to wider range of data types reveals need for pre-processing.
- Censor logic is applied to raw data before any regridding is done.
- May be specified separately for each forecast and/or observation field:
  - The **censor_thresh** entry is an array of thresholds.
  - The **censor_val** entry is an array of replacement values.
- Reflectivity Example:
  - Forecast reflectivity values are 0 or >= 35 dBZ.
  - Observed reflectivity values are continuous less than 35 dBZ.
  - Define observation censor:

```
censor_thresh = [ <35 ];
censor_val    = [ 0 ];
```

- Used to make the forecast and observation data more comparable.
- Can be used for range checking data.



Config: Climatology Data

- Required for anomaly correlation (**ANOM_CORR**), **SAL1L2**, and **VAL1L2**.
- NCEP monthly 2.5 degree
  - *match_day = FALSE*
- NCEP daily 1.0 degree
  - *match_day = TRUE*
- ERA daily 1.5 degree
  - *match_day = TRUE*
- Any other reference forecast.

```
climo_mean   = {

   file_name = [
    // List of file names
   ];
   field     = [
    // Same length as fcst.field
   ];
   regrid    = {
     method     = NEAREST;
     width      = 1;
     vld_thresh = 0.5;
   }
   time_interp_method = DW_MEAN;
   match_day          = FALSE;
   time_step          = 21600;
}
```

Config: Binned Climatologies

- Grid-Stat and Point-Stat process climatological distributions (i.e. climo mean and standard deviation).
- Binned climatologies affect only the computation of probabilistic statistics.
- Config file options:

```
climo_mean      = { … };   // Climo Mean Fields
climo_stdev     = { … };   // Climo Standard Deviation
climo_cdf_bins = 10;       // Number of Climo Bins
                           // Or array of bin values
```

- For each observation value, use the climo mean and standard deviation and compute a CDF value between 0 and 1.
- Place that observation into the correct climo CDF bin.
- Compute stats for all pairs within each bin.
- When climo mean and standard deviation are provided, derive the climatological probability values when computing Brier Skill Score.
- Bin number appended to the **VX_MASK** output column:
  - **FULL_BIN1, FULL_BIN2, …, FULL_BIN10**