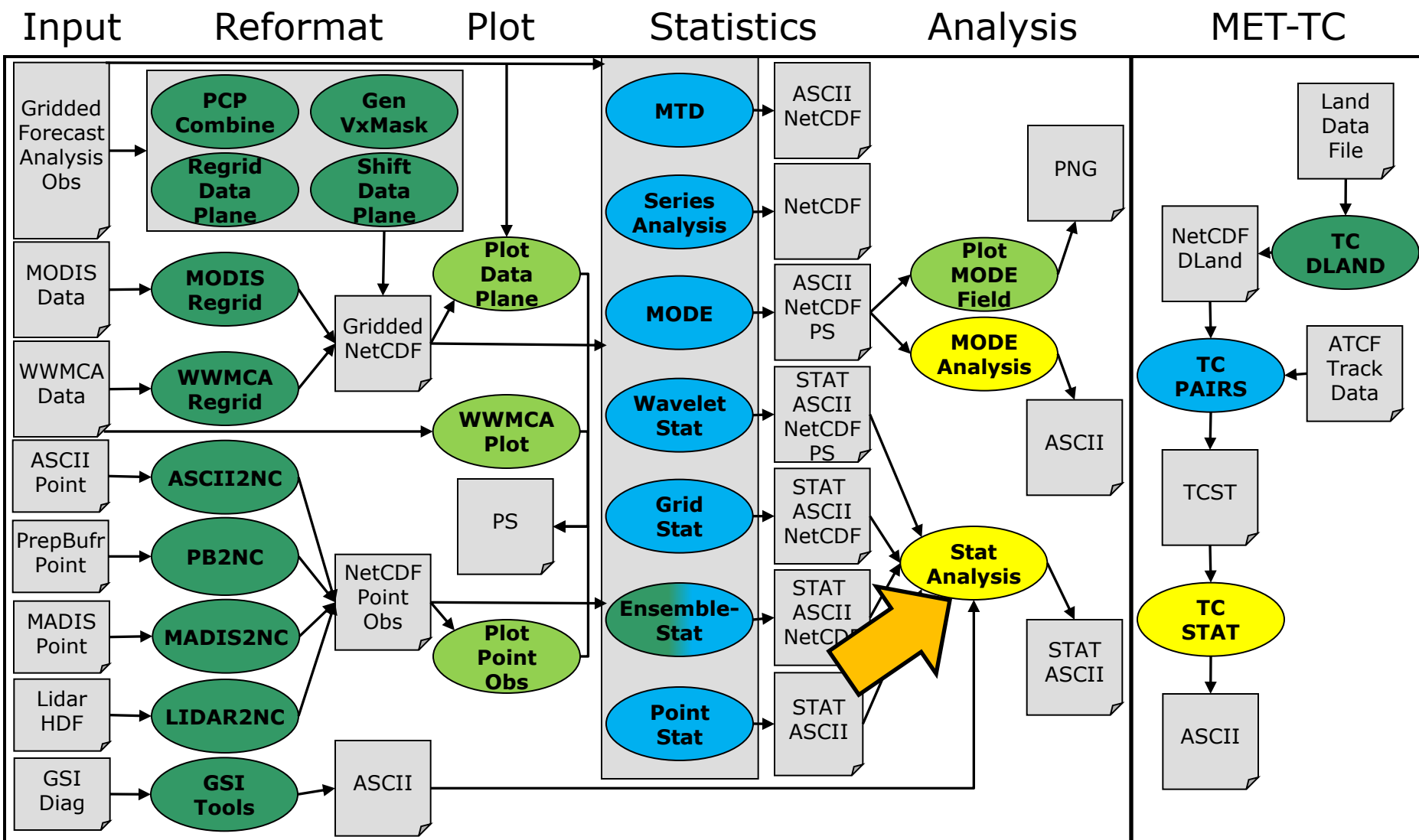


MET Stat-Analysis Tool

John Halley Gotway

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

Stat-Analysis Tool



Stat-Analysis: Overview

- **Filtering (filter)**
 - Filters out lines from one or more stat files
 - Filters based on user-specified filtering options
- **Summarizing (summary)**
 - Summary information from a single data column
 - Includes *mean, standard deviation, min, max, IQR, percentiles (10th, 25th, 50th, 75th, and 90th)*
- **Customized tool for AFWA (go_index)**
 - Computes GO Index, performance statistic used primarily by the US Air Force
 - Special case of the **ss_index** job type
- **Ramp**
 - Computes amount of change (delta) from one time to next
 - Delta thresholded to produce contingency table

Stat-Analysis: Overview

- **Aggregation**

- **aggregate** - aggregates stat data across multiple time steps or masking regions. ***Output line type is the same as input line type*** (i.e. SSVAR → SSVAR)
 - FHO, CTC, MCTC, SL1L2, SAL1L2, VL1L2, VAL1L2, PCT, NBRCNT, NBRCTC, GRAD, ISC, RHIST, PHIST, RELP, SSVAR
- **aggregate_stat** – aggregates across multiple times/regions then calculates statistics. Output line is different from input line types.

Valid line type combinations include:

-line_type FHO, CTC	yields -out_line_type CTS, ECLV
-line_type MCTC	yields -out_line_type MCTS
-line_type SL1L2, SAL1L2	yields -out_line_type CNT
-line_type VL1L2	yields -out_line_type VCNT
-line_type VL1L2, VAL1L2	yields -out_line_type WDIR
-line_type PCT	yields -out_line_type PSTD, PJC, PRC, ECLV
-line_type NBRCTC	yields -out_line_type NBRCTS
-line_type ORANK	yields -out_line_type RHIST, PHIST, RELP, SSVAR
-line_type MPR	yields -out_line_type FHO, CTC, CTS, MCTC, MCTS, CNT, SL1L2, SAL1L2, VL1L2, VCNT, PCT, PSTD, PJC, PRC, ECLV, WDIR

Stat-Analysis: Usage

Usage: stat_analysis

-lookin path

[-out filename]

[-tmp_dir path]

[-v level]

**[-config config_file
or -job as command
line options with
associated arguments**

- *filter*
- *summary*
- *aggregate*
- *aggregate_stat*
- *go_index*
- *ss_index*
- *ramp]*

-lookin	Path to .stat files – this can be: • a directory to be searched recursively for .stat files • one or more file names to be read • use one or more times
-out	Output name for ASCII file
-tmp_dir	Folder for temporary files
-v	Level of logging
-config	StatAnalysisConfig file

Stat-Analysis: Job Options

- Filter each of the 24 header columns:
 - Comma-separated list of strings (**model, desc, fcst/obs_var, fcst/obs_lev, obs_lev, obtype, vx_mask, line_type**)
 - Comma-separated list of times (**fcst/obs_lead, fcst/obs_init/valid_hour**)
 - Time bounds (**fcst/obs_init/valid_beg/end**)
 - Comma-separated list of thresholds (**fcst/obs/cov_thresh**)
- Filter data columns using **-column_str** to specify a column name and comma-separated list of strings to include:
 - *-line_type MPR -column_str OBS_SID KDEN,KBJC*
- Filter data columns using **-column_thresh** to specify a numeric column name and threshold:
 - *-line_type MPR -column_thresh FCST gt273*
- Use **-dump_row** to write lines used for a job to an output file.
- Use **-out_stat** to write full .stat lines with 22 header columns.
- Use **-set_hdr** to manually override output header columns.
- Many more job-specific options...

Stat-Analysis: Input/Output

- Input Files
 - MET .stat files
 - ASCII configuration file
- Output Files
 - MET .stat files

Stat-Analysis: Job Definition

- Defining jobs:
 1. Run a single job on the command line.
 - Many more job options are described in the MET User's Guide.
 - For example, **-out_thresh** defines the categorical threshold for:
-job aggregate_stat -line_type MPR -out_line_type CTC
 2. Run one or more jobs using a config file.
 - Everything above **jobs** entry defines filtering logic applied to all jobs.
 - STAT-Analysis reads input .stat files, applies top-level filtering logic, and *writes filtered output to a temporary file*.
 - Each job defined in the **jobs** entry reads that temporary file and applies any additional job-specific filtering criteria when running.

```
fcst_var = ["UGRD"];  
obs_var  = [];  
  
fcst_lev = [];  
obs_lev  = [];  
  
obtype = [];  
  
vx_mask = ["DTC165", "DTC166"];  
  
interp_mthd = ["DW_MEAN"];
```

```
jobs = [  
    "-job filter -line_type CTC -dump_row  
    outdir/job_filter_ctc_ugrd.stat",  
    "-job aggregate -line_type CTC -dump_row  
    outdir/job_aggregate_ctc_ugrd.stat"  
];
```

Stat-Analysis: Configuration

- Many configurable parameters – only set a few:
 - 10-m U-component of wind.
 - Aggregate stats over DTC165 and DTC166 regions
 - Accumulate only CTCs calculated using Distance-Weighted Mean interpolation
 - Dump lines included in accumulation
 - Dump aggregation to file

- OR -

can put it all in the
“jobs”area...

```
fcst_var = ["UGRD"];  
obs_var  = [];  
  
fcst_lev = [];  
obs_lev  = [];  
  
obtype = [];  
  
vx_mask = ["DTC165", "DTC166"];  
  
interp_mthd = ["DW_MEAN"];
```

```
jobs = [  
    "-job filter -line_type CTC -dump_row  
    outdir/job_filter_ctc_ugrd.stat",  
    "-job aggregate -line_type CTC -dump_row  
    outdir/job_aggregate_ctc_ugrd.stat"  
];
```

- OR -

```
jobs = [  
    "-job filter -line_type CTC -dump_row  
    out/job_filter_ctc_ugrd.stat",  
    "-job aggregate -line_type CTC -fcst_var UGRD  
    -vx_mask DTC165 -vx_mask DTC166  
    -interp_mthd DW_MEAN -dump_row  
    out/job_aggregate_ctc_ugrd.stat"  
];
```

Stat-Analysis: Run -job aggregate

```
stat_analysis -job aggregate -line_type CTC -fcst_var UGRD \
-vx_mask DTC165,DTC166 -interp_mthd DW_MEAN -dump_row \
out/job_aggregate.stat
```

Stat Analysis Filter Output in `job_aggregate.stat`

```
v8.0      WRF      360000
20070331_120000 20070331_120000
000000_20070331_103000_
20070331_133000 UGRD      Z10
UGRD      Z10      ADPSFC DTC165
DW_MEAN    9      >=5.000
>=5.000    NA      NA      CTC
934  32      43      32
827
```



	OBS			
F C S T		Y	N	
	Y	32	43	75
	N	32	827	859
		64	870	934

```
v8.0      WRF      360000
20070331_120000 20070331_120000
000000_20070331_103000_
20070331_133000 UGRD      Z10
UGRD      Z10      ADPSFC DTC166
DW_MEAN    9      >=5.000
>=5.000    NA      NA      CTC
2955 24      104      72
2755
```



	OBS			
F C S T		Y	N	
	Y	24	104	128
	N	72	2755	2827
		96	2859	2955

NOTE: Header modified to show only pertinent info

Stat-Analysis: Run –job aggregate

```
stat_analysis -job aggregate -line_type CTC -fcst_var UGRD \  
-vx_mask DTC165,DTC166 -interp_mthd DW_MEAN \  
-dump_row out/job_aggregate.stat
```

Stat Analysis Output in the file specified by –out flag (i.e. *stat_analysis.out*)

```
JOB_LIST:      -job aggregate  
      -fcst_var UGRD -vx_mask DTC165 -  
      vx_mask DTC166 -interp_mthd  
      DW_MEAN -line_type CTC -dump_row  
      out/aggregate2.stat  
COL_NAME: TOTAL FY_OY FY_ON FN_OY  
      FN_ON  
      CTC: 3889  56    147   104  
      3582
```

F C S T	OBS			
		Y	N	
	Y	56	147	251
	N	104	3582	317
		241	327	3889

Stat-Analysis: Run `-job aggregate_stat`

```
stat_analysis -job aggregate_stat -line_type CTC -out_line_type CTS \
-fcst_var UGRD -vx_mask DTC165,DTC166 -interp_mthd DW_MEAN \
-dump_row out/job_aggregate.stat
```

Aggregate_stat Output (*stat_analysis.out continued*)

```
COL_NAME: TOTAL BASER BASER_NCL BASER_NCU
BASER_BCL BASER_BCU FMEAN FMEAN_NCL FMEAN_NCU FMEAN_BCL
FMEAN_BCU ACC ACC_NCL ACC_NCU ACC_BCL ACC_BCU FBIAS
FBIAS_BCL FBIAS_BCU PODY PODY_NCL PODY_NCU PODY_BCL
PODY_BCU PODN PODN_NCL PODN_NCU PODN_BCL PODN_BCU POFD
POFD_NCL POFD_NCU POFD_BCL POFD_BCU FAR FAR_NCL
FAR_NCU FAR_BCL FAR_BCU CSI CSI_NCL CSI_NCU CSI_BCL
CSI_BCU GSS GSS_BCL GSS_BCU HK HK_NCL HK_NCU HK_BCL
HK_BCU HSS HSS_BCL HSS_BCU ODDS ODDS_NCL ODDS_NCU
ODDS_BCL ODDS_BCU CTS: 3889 0.04114 0.03534
0.04785 NA NA 0.05220 0.04564 0.05964 NA
NA 0.93546 0.92730 0.94276 NA NA 1.26875 NA
NA 0.35000 0.33516 0.36513 NA NA 0.96058
0.95400 0.96625 NA NA 0.03942 0.03375
0.04600 NA NA 0.72414 0.70987 0.73796 NA
NA 0.18241 0.17059 0.19486 NA NA 0.15955
NA NA 0.31058 0.23588 0.38528 NA NA 0.27519
NA NA 13.12088 9.11454 18.88823 NA NA
```

		OBS		
F C S T		Y	N	
	Y	56	147	203
	N	104	3582	3686
		160	3729	3889

Base Rate: 0.04
Freq Bias: 1.27
PODY: 0.35
FAR: 0.72
CSI: 0.18
GSS: 0.15

Stat-Analysis: Run –job summary

stat_analysis -job summary -fcst_var UGRD -interp_mthd DW_MEAN \
-line_type CTS -column GSS -dump_row out/job_summary.stat

#	Description	
1	Column Name	Summary
2	Total	3
3-7	Mean* <i>Includes normal and bootstrap upper and lower confidence limits</i>	0.109
8-10	Standard deviation** <i>Includes bootstrap upper and lower confidence limits</i>	0.150
11	Minimum value	-0.036
12	10 th percentile	-0.008
13	25 th percentile	0.032
14	Median (50 th percentile)	0.101
15	75 th percentile	0.182
16	90 th percentile	0.231
17	Maximum value	0.263

Summary Output (stat_analysis.out cont.)

```
COL_NAME: TOTAL MEAN
MEAN_NCL MEAN_NCU MEAN_BCL
MEAN_BCU STDEV      STDEV_BCL
STDEV_BCU MIN        P10
P25      P50      P75      P90
MAX
SUMMARY: 3      0.10963
-0.26321 0.48247 -0.03613
0.26370  0.15009 0.00000
0.17311  -0.03613 -0.00864
0.03259 0.10131 0.18251
0.23122 0.26370
```

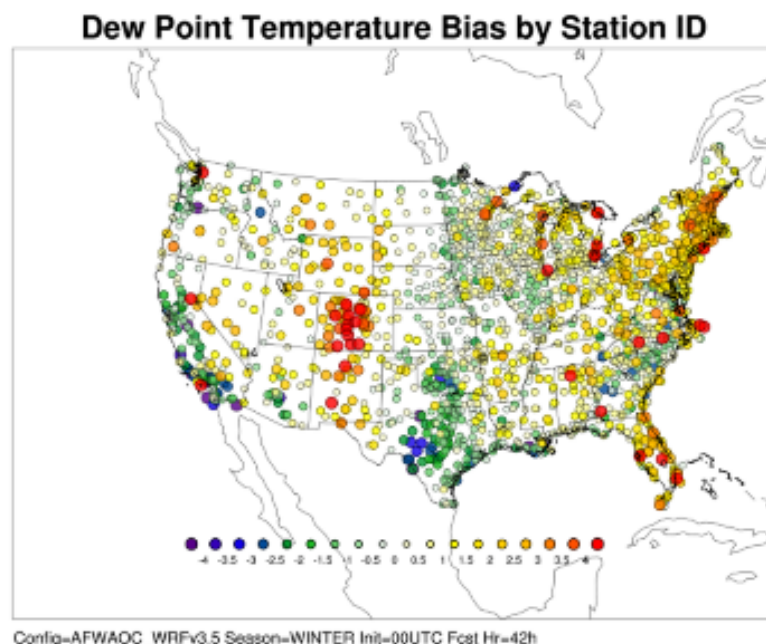
Stat-Analysis: Run option “-by”

-by option is very powerful

- Runs a **single job** but writes separate output for **each unique combination** of the values found in the header column(s) specified.

```
stat_analysis \  
-lookin out/grid_stat \  
-job aggregate_stat \  
-line_type SL1L2 -out_line_type CNT \  
-out_stat cnt.txt \  
-by FCST_VAR,FCST_LEV,VX_MASK
```

```
stat_analysis \  
-lookin out/point_stat \  
-job aggregate_stat \  
-line_type MPR -out_line_type CNT \  
-out_stat cnt_by_SID.txt \  
-by OBS_SID
```



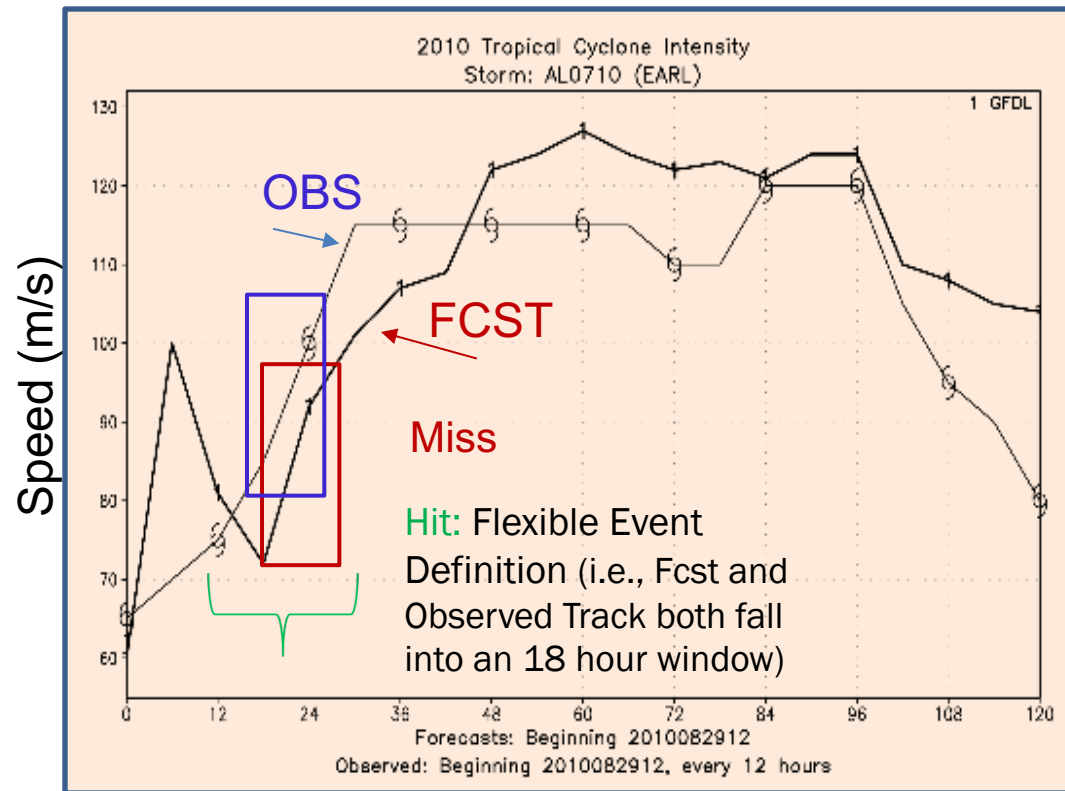
Stat-Analysis: Run –job ramp

- Original support for tropical cyclone verification then expanded for other applications

Events identified by
automated algorithm or
user defined criteria of
 DI / DT

*Also applicable to any
intensity change event*

- *Renewable Energy Ramps*
- *Solar Wind during CMEs*
- *Magenetic Field*
- *Temperature*



Categorical statistics for RI or Rapid Weakening events (RW) can be computed

Stat-Analysis: Run -job ramp

Up-ramp

```
stat_analysis -lookin /path/to/file \  
-job ramp -ramp_type dydt \  
-ramp_thresh gt200 \  
-ramp_time 0030 \  
-by OBS_SID,VX_MASK,MODEL \  
-out_line_type CTC,CTS,MPR \  
-out ramp_cts/20150324/upramp_gt200.txt \  
-v 4 -log ramp_cts/20150324/upramp_gt200.log
```

Assumes input is MPR line type

- Find Increase in intensity
- Greater than 200 W/m²
- In a 30 minute period
- Ramp must occur at same time in both forecast and obs for it to be a hit

Down-ramp

```
bin/stat_analysis -lookin /path/to/file \  
-job ramp -ramp_type dydt \  
-ramp_thresh lt-200 \  
-ramp_time 0030 -ramp_exact false -ramp_width 02 \  
-by OBS_SID,VX_MASK,MODEL \  
-out_line_type CTC,CTS,MPR \  
-out ramp_cts/20150324/downramp_lt-200.txt \  
-v 4 -log ramp_cts/20150324/downramp_lt-200.log
```

- Find Decrease in intensity
- Greater than 200 W/m²
- In a 30 minute period
- Ramp must occur within 2 hours in both forecast and obs for it to be a hit

Stat-Analysis: Wind Direction Example

Point-Stat: VL1L2 Lines

VX MASK	THRESH	LINE_TYPE	TOTAL	UFBAR	VFBAR	UOBAR	VOBAR	UVFOBAR	UVFFBAR	UVOOBAR
DTC_165	>=1.000	VL1L2	653	1.91117	0.07900	1.40658	-0.06126	13.01039	18.12575	20.31649
DTC_165	>=3.000	VL1L2	279	3.13561	-0.35096	2.87061	-0.30072	26.50472	30.03257	38.25362
DTC_165	>=5.000	VL1L2	96	5.21268	-2.74580	5.47813	-2.01667	49.90791	51.10427	70.78802
DTC_166	>=1.000	VL1L2	2431	-1.62742	0.25391	-1.23402	-0.04393	18.48309	29.70179	21.89615
DTC_166	>=3.000	VL1L2	1610	-1.84581	0.16061	-1.47491	-0.11217	24.45214	36.67400	29.36032
DTC_166	>=5.000	VL1L2	520	-0.93518	-0.45435	-0.25923	-0.49558	37.21821	52.51917	47.26483

Stat-Analysis: aggregate_stat jobs

```
JOB_LIST:      -job aggregate_stat -fcst_thresh >=1.000 -line_type VL1L2 -out_line_type WDIR
COL_NAME: TOTAL FBAR      OBAR      ME      MAE
ROW_MEAN_WDIR: 2      183.25038 0.22749  -3.02289  7.88372
AGGR_WDIR: 3084  103.87238 85.96574 -17.90663 NA
```

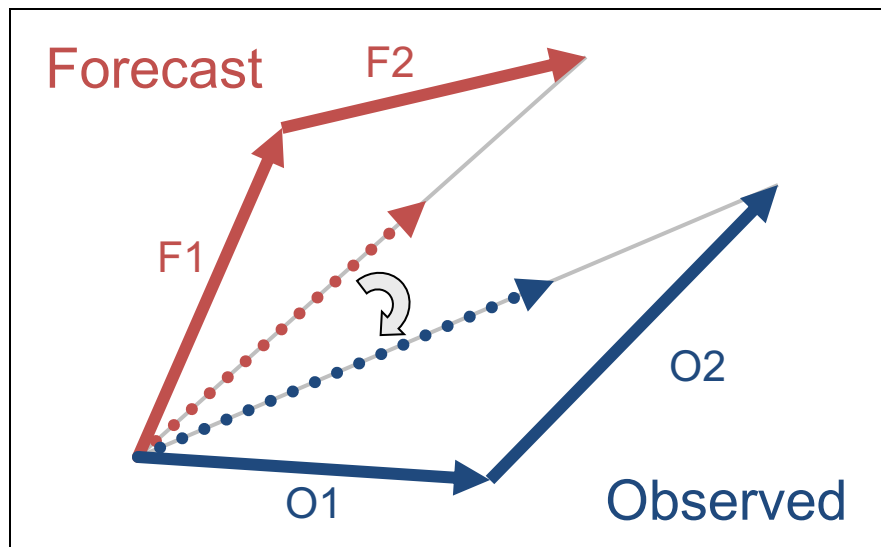
```
JOB_LIST:      -job aggregate_stat -fcst_thresh >=3.000 -line_type VL1L2 -out_line_type WDIR
COL_NAME: TOTAL FBAR      OBAR      ME      MAE
ROW_MEAN_WDIR: 2      5.67967 0.81565  -4.86402  4.86402
AGGR_WDIR: 1889  94.38140 80.45939 -13.92200 NA
```

```
JOB_LIST:      -job aggregate_stat -fcst_thresh >=5.000 -line_type VL1L2 -out_line_type WDIR
COL_NAME: TOTAL FBAR      OBAR      ME      MAE
ROW_MEAN_WDIR: 2      0.93288 338.91179 -22.02109 22.02109
AGGR_WDIR: 616  358.38152 319.08761 -39.29391 NA
```

Stat-Analysis: Wind Direction Example

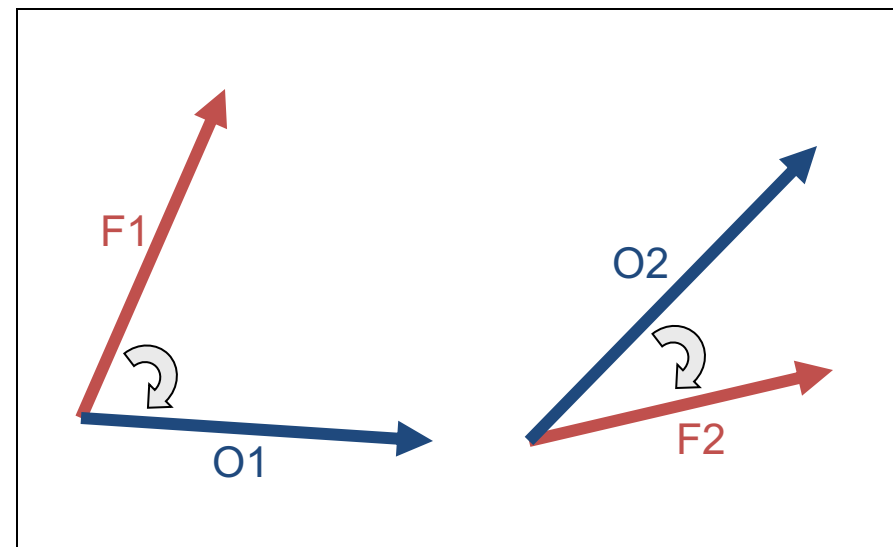
- **AGGR_WDIR**

1. Aggregate **VL1L2** partial sums lines
2. Derive wind directions and errors



- **ROW_MEAN_WDIR**

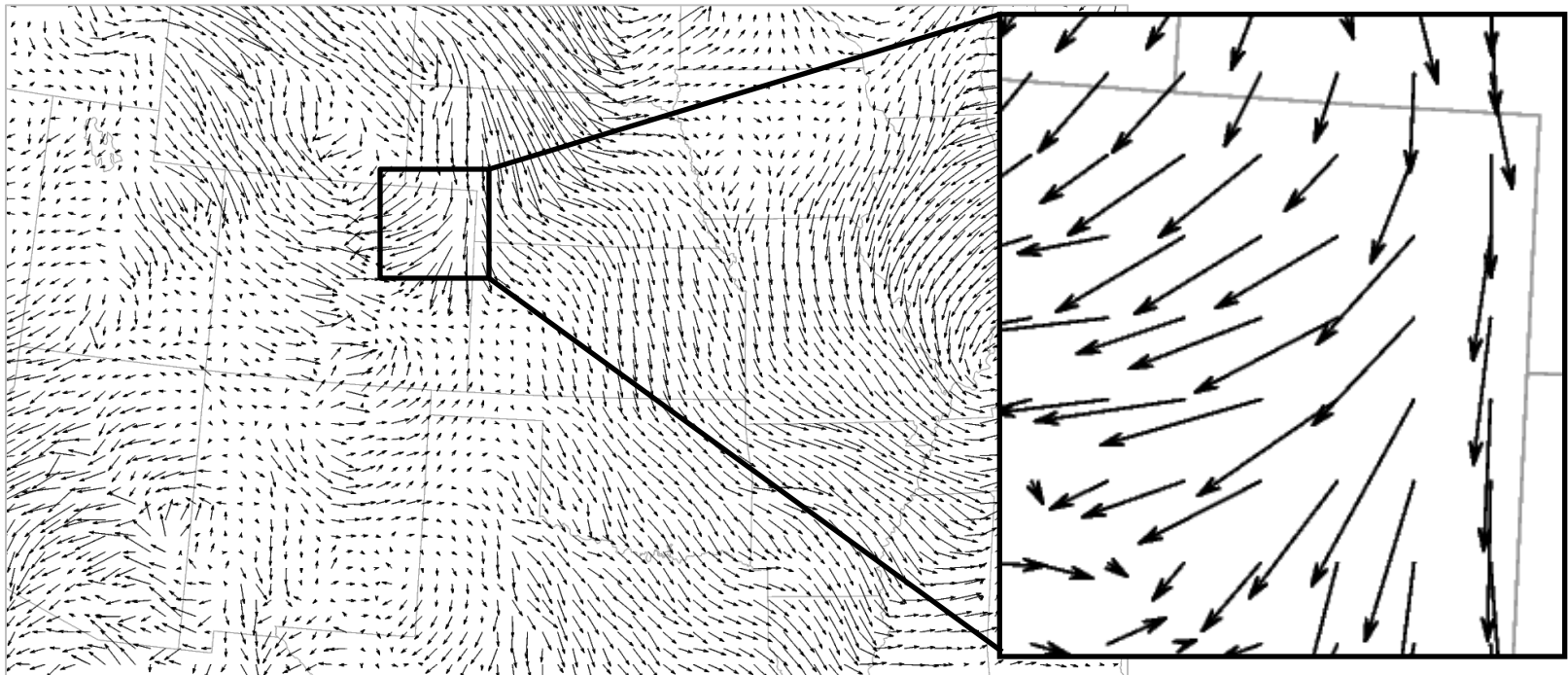
1. Derive wind directions and errors for each **VL1L2** line
2. Compute mean of errors



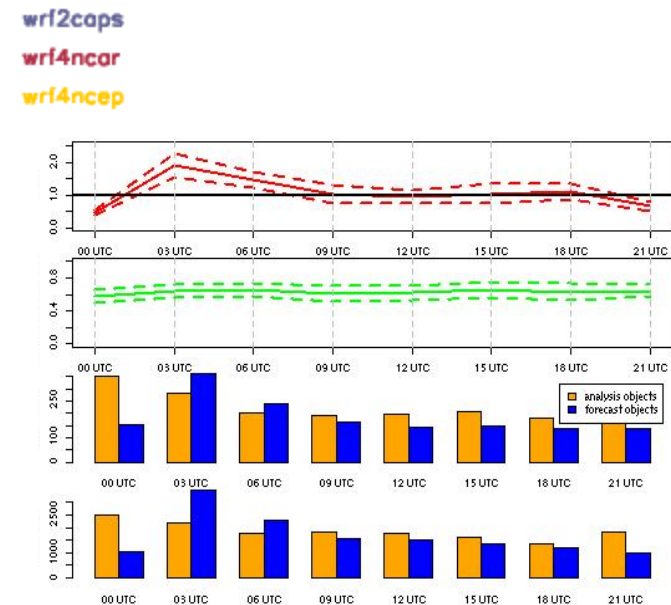
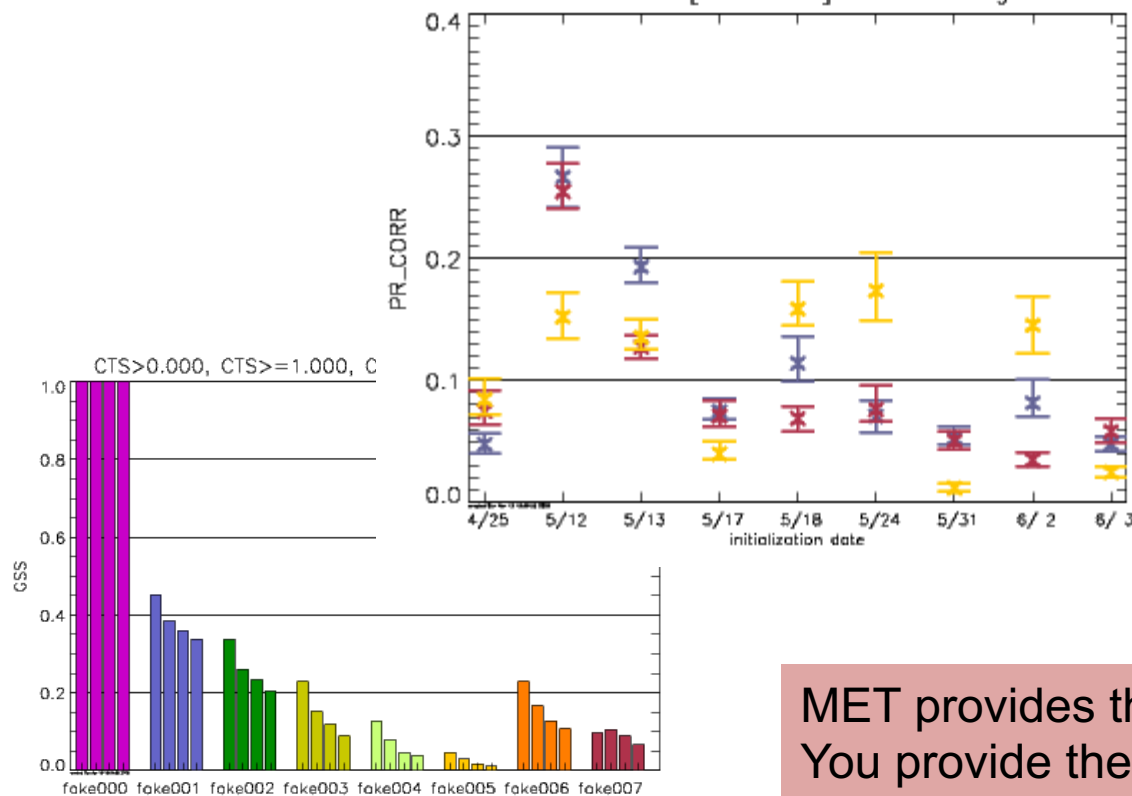
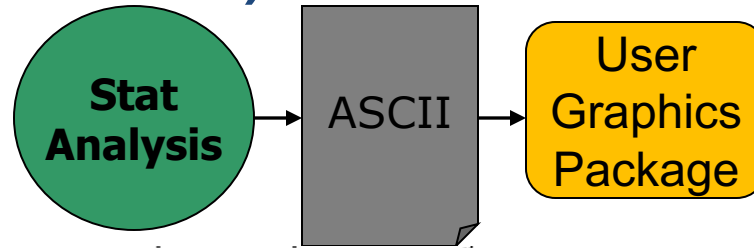
Stat-Analysis: Wind Direction Example

Reminder: When aggregating, wind directions can cancel out. We suggest you

- Verify over regions with unimodal wind direction.
- Verify u and v components separately.



Use your favorite plotting software (Or METviewer)



MET provides the analysis in ascii output.
You provide the graphing / plotting capability.