Stat-Analysis Tool

- Filtering
- Summarizing
- Aggregating

of Grid-Stat, Point-Stat,
Ensemble-Stat & Wavelet-Stat output

Tara Jensen
What can Stat Analysis do?

Questions to MET Help - Can I get...

Q: Overall statistics for all gridded observations compared to the forecasts for hours 0 through 24 together?
A: Yes - using Stat Analysis Tool on Grid-Stat output

Q: My contingency table statistics aggregated over multiple runs?
A: Yes – using Stat Analysis Tool on any output

Q: Long-term statistics at individual sites (e.g., mean absolute error or RMS error for daily forecasts for a month)?
A: Yes - using Stat Analysis Tool on Point-Stat output

Q: Statistics aggregated for a large number (N) of individual stations in one simultaneous run?
A: Yes – but it would be cumbersome. You would have to configure Stat Analysis Tool to run (N) number of jobs OR use our soon to be released METViewer tool.

Copyright 2015, University Corporation for Atmospheric Research, all rights reserved
For Stat Analysis Tool:
MET provides the analysis in ASCII output. You provide the graphing / plotting capability.
Stat Analysis Jobs

- **Filtering**
  - *filter* - filters out lines from one or more stat files based on user-specified filtering options.

- **Summarizing**
  - *summary* - produces summary information from a single column of data including:
    
    mean, standard deviation, min, max, and the 10th, 25th, 50th, 75th, and 90th percentiles.

- **Customized tool for AFWA**
  - *go_index* - computes the GO Index, a performance statistic used primarily by the United States Air Force.
Stat Analysis Jobs

- **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)
  
  - `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
- `line_type` MCTC yields `out_line_type` MCTS
- `line_type` SL1L2, SAL1L2 yields `out_line_type` CNT
- `line_type` VL1L2, VAL1L2 yields `out_line_type` WDIR
- `line_type` PCT yields `out_line_type` PSTD, PJC, PRC
- `line_type` NBRCTC yields `out_line_type` NBRCTS
- `line_type` MPR yields `out_line_type` FHO, CTC, CTS MCTC, MCTS, CNT, SL1L2 SAL1L2, PCT, PSTD, PJC, PRC

Copyright 2015, University Corporation for Atmospheric Research, all rights reserved
**Stat Analysis Tool: Usage**

Usage: `stat_analysis`

- `lookin path`  
  [-out filename]  
  [-tmp_dir path]  
  [-v level]

- `config config_file`

or `--job` at command line

options with associated arguments

- `filter`
  [summary]
  [aggregate]
  [aggregate_stat]
  [go_index]

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>-lookin</code></td>
<td>Path to *.stat files – this can be a directory or a single file name</td>
</tr>
<tr>
<td><code>-out</code></td>
<td>Output name for ASCII file</td>
</tr>
<tr>
<td><code>-tmp_dir</code></td>
<td>Folder for temporary files</td>
</tr>
<tr>
<td><code>-v</code></td>
<td>Level of logging</td>
</tr>
<tr>
<td><code>-config</code></td>
<td>StatAnalysisConfig file</td>
</tr>
<tr>
<td><code>filter</code></td>
<td>See previous 2 slides</td>
</tr>
<tr>
<td><code>summary</code></td>
<td>See previous 2 slides</td>
</tr>
<tr>
<td><code>aggregate</code></td>
<td>See previous 2 slides</td>
</tr>
<tr>
<td><code>aggregate_stat</code></td>
<td>See previous 2 slides</td>
</tr>
<tr>
<td><code>go_index</code></td>
<td>See previous 2 slides</td>
</tr>
</tbody>
</table>

Copyright 2015, University Corporation for Atmospheric Research, all rights reserved
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 -

You can put it all in the “jobs” area...

```bash
fcst_var = ["UGRD"];
obs_var = [];
fcst_lev = [];
obsv_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
    outdir/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
    outdir/job_aggregate_ctc_ugrd.stat"
];
```
Stat Analysis Tool: Run `--job aggregate`

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

Stat Analysis Filter Output in `job_aggregate.stat`

<table>
<thead>
<tr>
<th>V4.1</th>
<th>WRF</th>
<th>360000</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>20070331_120000 20070331_120000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>20070331_133000 UGRD Z10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>UGRD Z10 ADPSFC DTC165</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DW_MEAN 9 &gt;=5.000 &gt;=5.000 NA NA CTC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>934 32 43 32</td>
</tr>
<tr>
<td></td>
<td></td>
<td>827</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>V4.1</th>
<th>WRF</th>
<th>360000</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>20070331_120000 20070331_120000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>20070331_133000 UGRD Z10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>UGRD Z10 ADPSFC DTC166</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DW_MEAN 9 &gt;=5.000 &gt;=5.000 NA NA CTC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2955 24 104 72</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2755</td>
</tr>
</tbody>
</table>

(Note: header modified to show only pertinent info)
Stat Analysis Tool: Run `-job aggregate`

```
"-job aggregate -line_type CTC -fcst_var UGRD -vx_mask DTC165 -vx_mask 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`)

<table>
<thead>
<tr>
<th>COL_NAME: TOTAL FY_OY FY_ON FN_OY FN_ON</th>
</tr>
</thead>
<tbody>
<tr>
<td>CTC: 3889 56 147 104 3582</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>OBS</th>
</tr>
</thead>
<tbody>
<tr>
<td>FCST</td>
</tr>
<tr>
<td>----</td>
</tr>
<tr>
<td>Y</td>
</tr>
<tr>
<td>N</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
Stat Analysis Tool: Run `job aggregate_stat`

"-job aggregate_stat -line_type CTC -out_line_type CTS -fcst_var UGRD -vx_mask DTC165 -vx_mask DTC166 -interp_mthd DW_MEAN -dump_row out/job_aggregate_stat.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

<table>
<thead>
<tr>
<th>OBS</th>
<th>F CST</th>
<th>Y</th>
<th>N</th>
<th>N</th>
<th>203</th>
<th>1.27</th>
<th>0.35</th>
<th>0.72</th>
<th>0.18</th>
<th>0.15</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>56</td>
<td>147</td>
<td>203</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y</td>
<td></td>
<td>104</td>
<td>3582</td>
<td>3686</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>160</td>
<td>3729</td>
<td>3889</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

Copyright 2015, University Corporation for Atmospheric Research, all rights reserved
Stat Analysis Tool: Run –job summary

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

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

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
```
Use your favorite plotting software

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

Copyright 2015, University Corporation for Atmospheric Research, all rights reserved
Please feel free to send your contributions to met_help@ucar.edu