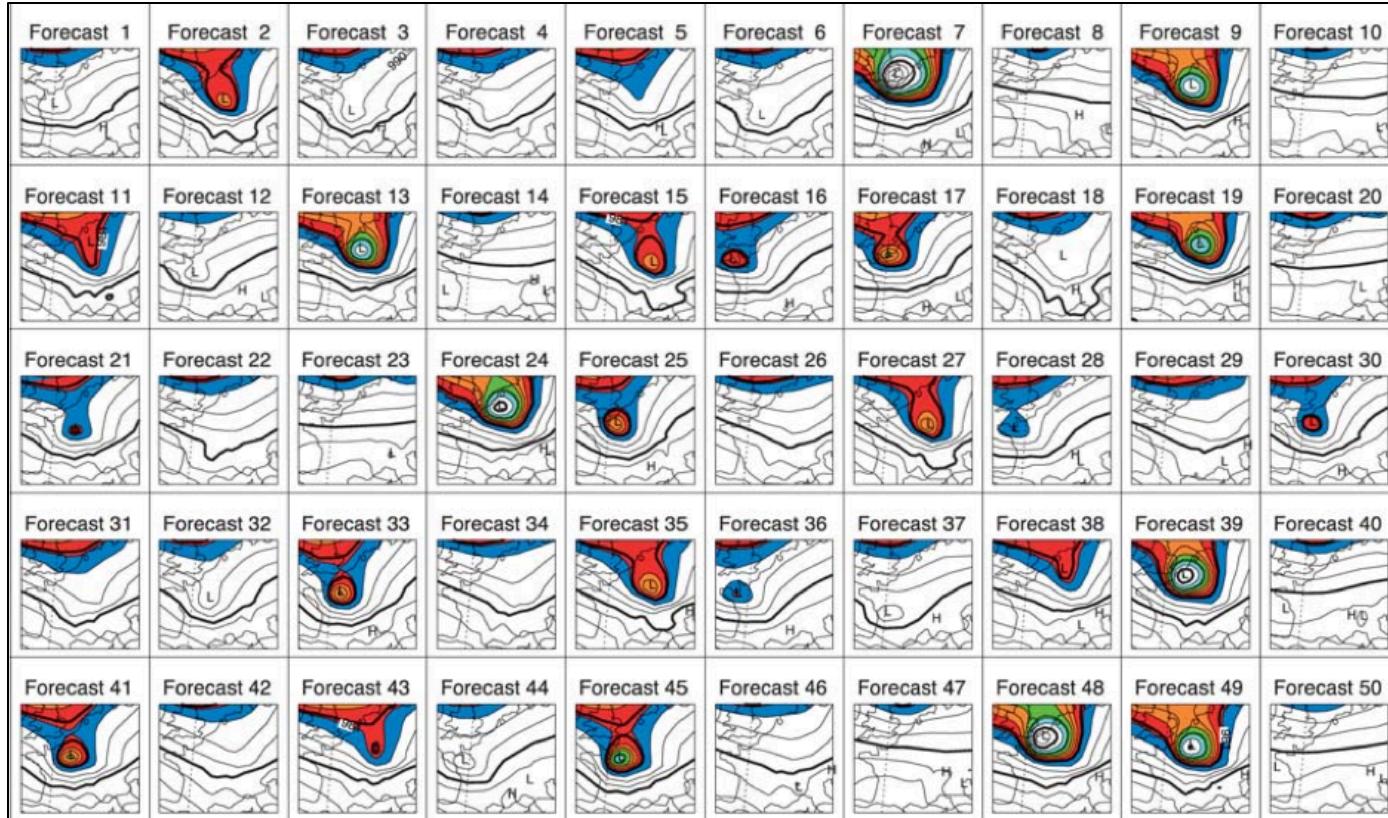


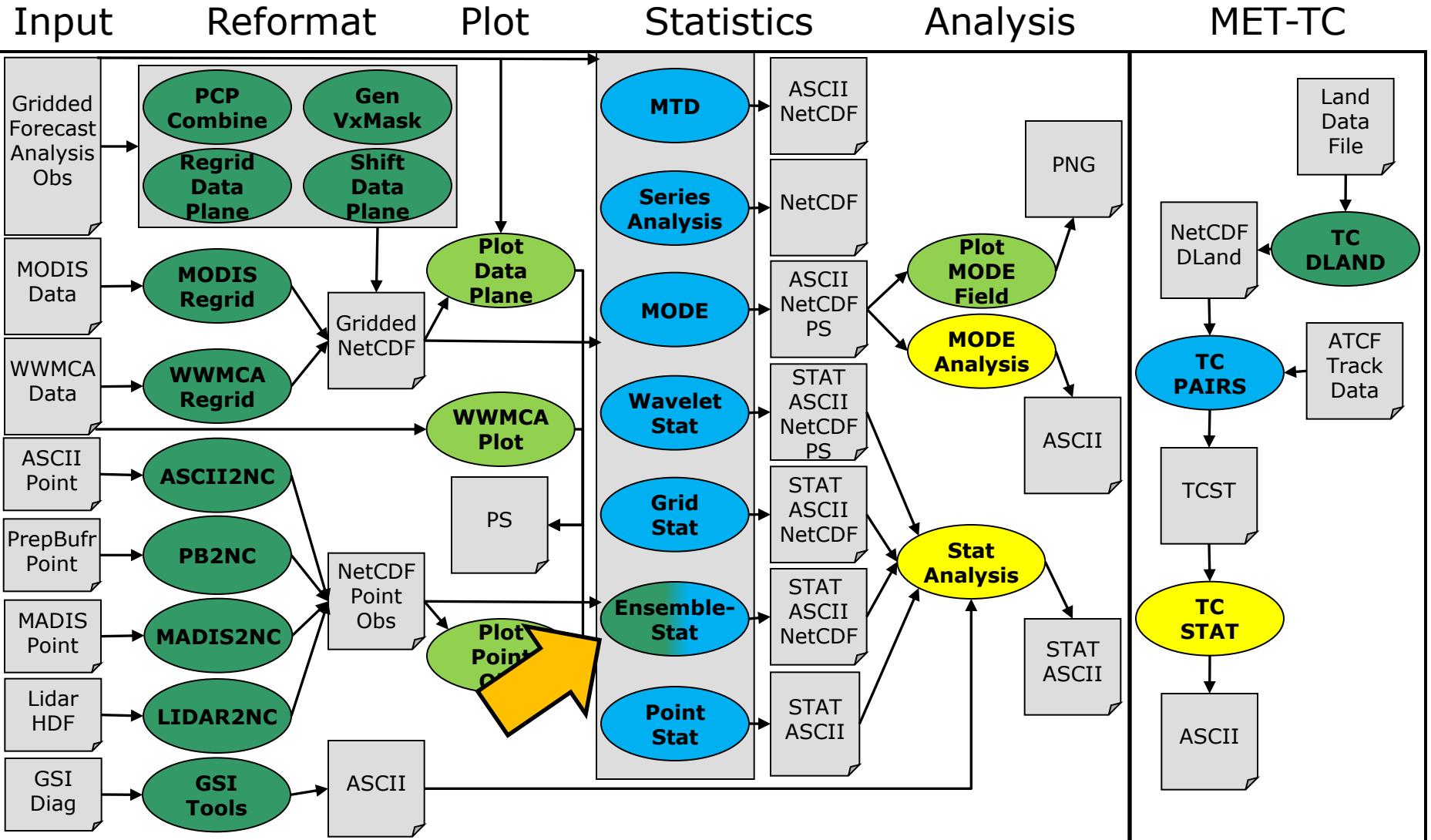
Ensemble_Stat



Presenter: Tara Jensen

copyright 2018, UCAR, all rights reserved

Ensemble-Stat Tool



Ensemble-Stat Capabilities

Reads:

- Gridded ensemble member files
- *Gridded AND point* observations files

Calculates:

- Ensemble Mean, Standard Deviations, Mean \pm 1 SD fields
- Ensemble Min, Max, and Range fields
- Ensemble Valid Data Count field
- Ensemble Relative Frequency by threshold fields (i.e. probability)
- Rank and PIT Histograms (if Obs Field Provided)
- Ensemble Spread-Skill (if Obs Field Provided)
- Ensemble continuous statistics (i.e. RMSE, ME, if obs)

Writes:

- Stat file with Continuous Statistics, Rank Histogram, PIT Histogram, Spread-Skill partial sums, and Point Observation Ranks
- Gridded field of Observation Ranks to a NetCDF file

Ensemble Stat Tool: Usage

Usage: ensemble_stat

```
n_ens ens_file_1 \
... ens_file_n OR
ens_file_list OR *
config_file
[-grid_obs file]
[-point_obs file]
[-ens_mean_file]
[-obs_valid_beg time]
[-obs_valid_end time]
[-outdir path]
[-log file]
[-v level]
[-compress level]
```

Number of Ensemble members followed by list of ensemble member names OR ens_file_list (the name of an ASCII file with names of members)
Config file name
Name of gridded or point observed file – Required if Rank Histograms desired (optional)
Specify model mean file (override sample ens mean for computation of ECN, SSVAR, ORANK line types)
YYYYMMDD[HH[MMSS]] format to set the beginning and end of the matching observation time window (optional)
Set output directory (optional)
Outputs log messages to the specified file (optional)
Set level of verbosity (optional)
Sets the compressions level for storing NetCDF data

Ensemble-Stat: Configuration

- Many configurable parameters
 - ens = fields to summarize
 - ens_thresh - All members must be available
 - vld_thresh – all data in grid must be valid
 - 24hr Accumulated Precip (APCP)
 - Composite Reflectivity (REFC)
 - N-S component of Wind (UGRD)
- Thresholds used for Ensemble Relative Freq (i.e. probability)
- GRIB1_ptv = 129; Use GRIB Table 129 instead of Table 2

```
//  
// Ensemble product fields to be  
// processed  
// (i.e. mean, min, max, stdev fields)  
//  
ens = {  
    ens_thresh = 1.0;  
    vld_thresh = 1.0;  
  
    field = [  
        {  
            name      = "APCP";  
            level     = [ "A24" ];  
            cat_thresh = [ >0.0, >=10.0 ];  
        },  
        {  
            name      = "REFC";  
            level     = [ "L0" ];  
            cat_thresh = [ >=35.0 ];  
            GRIB1_ptv = 129;  
        },  
        {  
            name      = "UGRD";  
            level     = [ "Z10" ];  
            cat_thresh = [ >=5.0 ];  
        },  
    ];  
}
```

Ensemble-Stat: Configuration

- Many configurable parameters – only set a few:
 - Fcst specifies fields to be verified
 - ADPSFC message type for point obs
 - 24hr precip for gridded obs field
 - Bin size for spread-skill calculation is 0.1 mm
 - Bin size for probability integral transform statistics is 0.05 mm

```
// Forecast and observation fields to be
// verified (i.e. RHIST, PHIST, SSVAR)
//
fcst = {
    field = [
        {
            name           = "APCP";
            level          = [ "A24" ];
        }
    ];
}
obs = fcst;
```

```
// Point observation filtering options
// May be set separately in each "obs.field" entry
//
message_type      = [ "ADPSFC" ];
sid_exc          = [];
obs_thresh       = [ NA ];
obs_quality     = [];
duplicate_flag   = NONE;
obs_summary      = NONE;
obs_perc_value  = 50;
skip_const       = FALSE;

//
// Ensemble bin sizes
// May be set separately in each "obs.field" entry
//
ens_ssvar_bin_size = 0.1;
ens_phist_bin_size = 0.05;
```

Ensemble-Stat Tool: Run

```
ensemble_stat \
6 sample_fcst/2009123112/*gep*/d01_2009123112_02400.grib \
config/EnsembleStatConfig \
-grid_obs sample_obs/ST4/ST4.2010010112.24h \
-point_obs out/ascii2nc/precip24_2010010112.nc \
-outdir out/ensemble_stat -v 2
```

NOTE:

You can pass in a path with wildcards to pull out the files you would like to process or you can pass in an ASCII filename that contains a list of ensemble members

Gridded and Obs fields are included for use in calculating Rank Histogram, PIT Histogram, and Spread-Skill

Ensemble Stat Tool: Run

```
*** Running Ensemble-Stat on APCP using GRIB forecasts, point observations, and gridded observations ***
DEBUG 1: Default Config File: /d3/projects/MET/MET_releases/met-6.0/data/config/EnsembleStatConfig_default
DEBUG 1: User Config File: config/EnsembleStatConfig
GSL_RNG_TYPE=mt19937
GSL_RNG_SEED=1
DEBUG 1: Ensemble Files[6]:
DEBUG 1: ./data/sample_fcst/2009123112/arw-fer-gep1/d01_2009123112_02400.grib
DEBUG 1: ./data/sample_fcst/2009123112/arw-fer-gep5/d01_2009123112_02400.grib
DEBUG 1: ./data/sample_fcst/2009123112/arw-sch-gep2/d01_2009123112_02400.grib
DEBUG 1: ./data/sample_fcst/2009123112/arw-sch-gep6/d01_2009123112_02400.grib
DEBUG 1: ./data/sample_fcst/2009123112/arw-tom-gep3/d01_2009123112_02400.grib
DEBUG 1: ./data/sample_fcst/2009123112/arw-tom-gep7/d01_2009123112_02400.grib
DEBUG 1: Gridded Observation Files[1]:
DEBUG 1: ./data/sample_obs/ST4/ST4.2010010112.24h
DEBUG 1: Point Observation Files[1]:
DEBUG 1: ./out/ascii2nc/precip24_2010010112.nc
DEBUG 2:
DEBUG 2: -----
DEBUG 2: -----
DEBUG 2: Processing ensemble field: APCP/A24
DEBUG 2:
DEBUG 2: -----
...
Processing gridded verification APCP_24/A24 versus APCP_24/A24, for observation type MC_PCP, over region FULL, for
interpolation method UW_MEAN(1), using 15480 pairs.
-----
DEBUG 1: Output file: out/ensemble_stat/ensemble_stat_20100101_120000V.stat
DEBUG 1: Output file: out/ensemble_stat/ensemble_stat_20100101_120000V_rhist.txt
DEBUG 1: Output file: out/ensemble_stat/ensemble_stat_20100101_120000V_phist.txt
DEBUG 1: Output file: out/ensemble_stat/ensemble_stat_20100101_120000V_orank.txt
DEBUG 1: Output file: out/ensemble_stat/ensemble_stat_20100101_120000V_ssvar.txt
DEBUG 1: Output file: out/ensemble_stat/ensemble_stat_20100101_120000V_ens.nc
DEBUG 1: Output file: out/ensemble_stat/ensemble_stat_20100101_120000V_orank.nc
```

Ensemble Stat Tool: txt Output

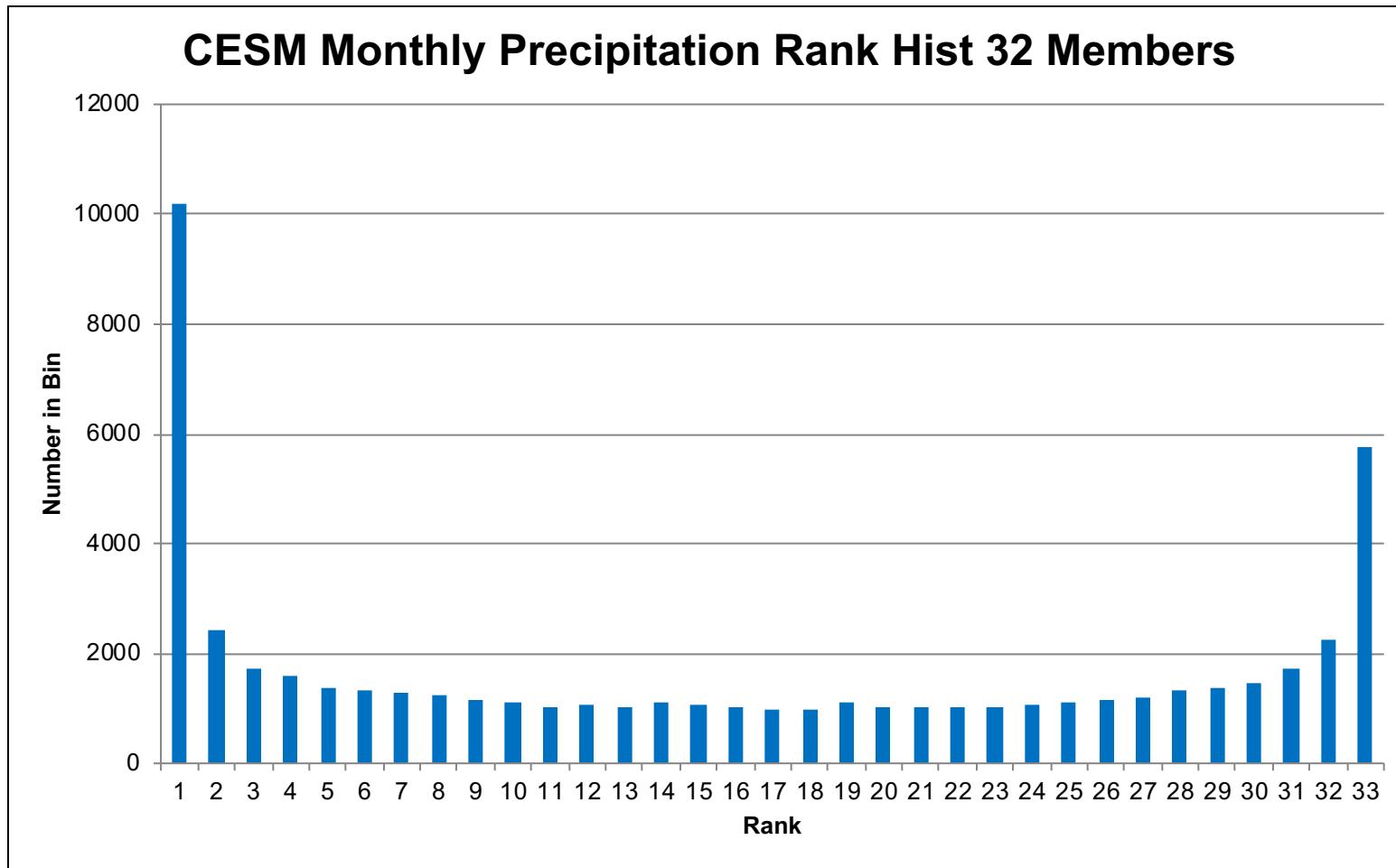
*Output from *_rhist.txt*

VERSION MODEL	FCST LEAD	FCST VALID BEG	FCST VALID END	OBS LEAD	OBS VALID BEG	OBS VALID END	FCST VAR	FCST LEV	OBS VAR	OBS LEV	OBTYPE	VX_MASK	INTERP_MTHD	INTERP_PNTS	FCST_THRESH	OBS_THRESH	COV_THRESH	ALPHA	LINE_TYPE	TOTAL	CRPS	IGN	N_RANK	RANK_1	RANK_2	RANK_3	RANK_4	RANK_5	RANK_6	RANK_7				
VX.Y	WRF	240000	20100101_120000	20100101_120000	000000	20100101_103000	20100101_133000	APCP_24	A24	APCP_24	A24	ADPSFC	FULL	UW_MEAN	1	NA	NA	NA	NA	RHIST	1125	8.21904	6.53721	7	261	160	138	141	149	111	165	CRPS	IGN	RANK HIST

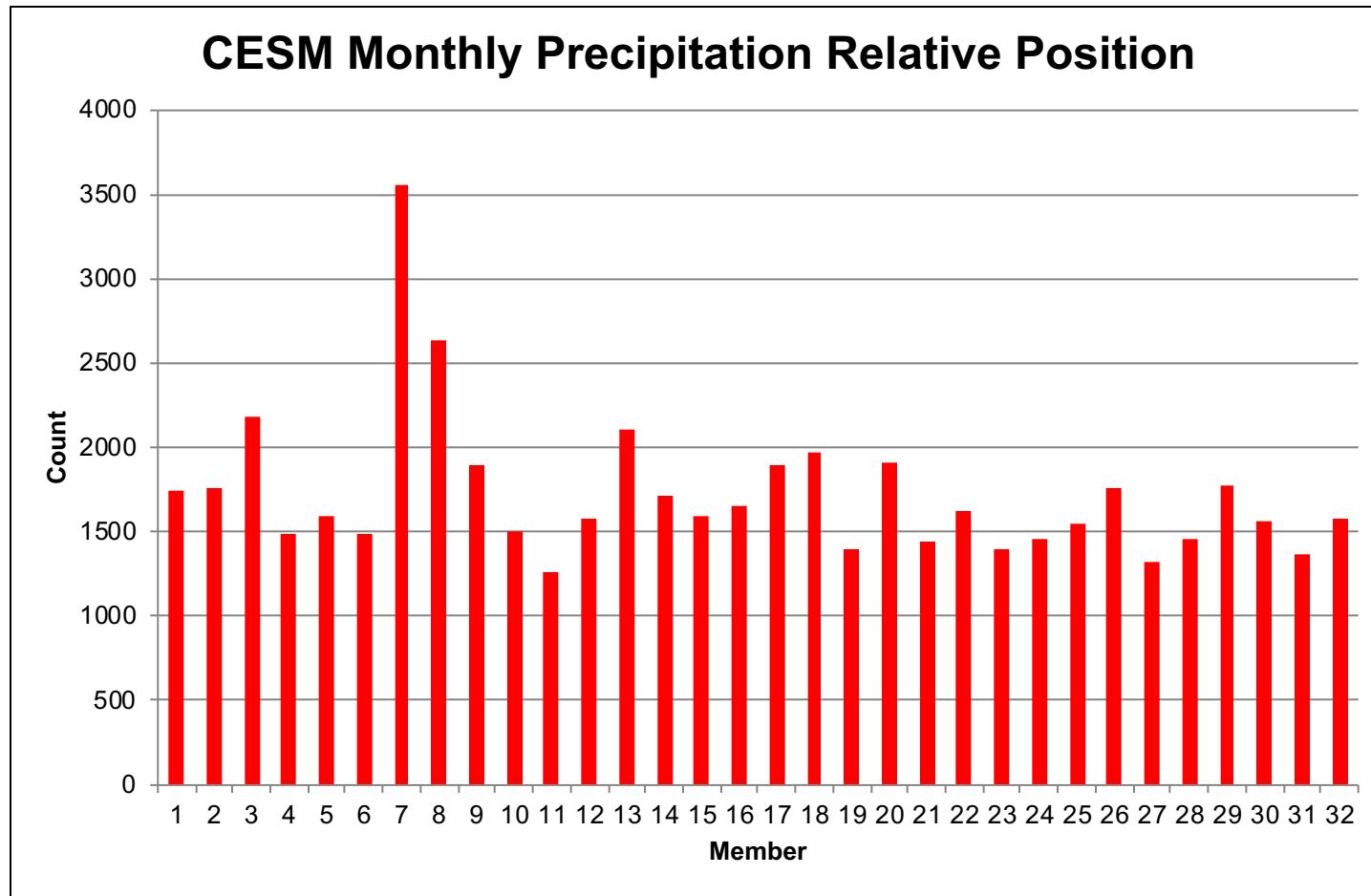
*Output from *_phist.txt*

VERSION MODEL	FCST LEAD	FCST VALID BEG	FCST VALID END	OBS LEAD	OBS VALID BEG	OBS VALID END	FCST VAR	FCST LEV	OBS VAR	OBS LEV	OBTYPE	VX_MASK	INTERP_MTHD	INTERP_PNTS	FCST_THRESH	OBS_THRESH	COV_THRESH	ALPHA	LINE_TYPE	TOTAL	BIN_SIZE	N_BIN	BIN_1	BIN_2	BIN_3	BIN_4	BIN_5	BIN_6	BIN_7	BIN_8	BIN_9	BIN_10	BIN_11	BIN_12	BIN_13	BIN_14	BIN_15	BIN_16	BIN_17	BIN_18	BIN_19	BIN_20	VX.Y	WRF	240000	20100101_120000	20100101_120000	000000	20100101_103000	20100101_133000	APCP_24	A24	APCP_24	A24	ADPSFC	FULL	UW_MEAN	1	NA	NA	NA	NA	PHIST	55296	0.05	20	9261	3135	2565	2258	2237	2043	2084	2167	2059	2155	2205	2202	2198	2174	2097	2153	1992	2023	2315	5973	Probability integral transform histogram
---------------	-----------	----------------	----------------	----------	---------------	---------------	----------	----------	---------	---------	--------	---------	-------------	-------------	-------------	------------	------------	-------	-----------	-------	----------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	------	-----	--------	-----------------	-----------------	--------	-----------------	-----------------	---------	-----	---------	-----	--------	------	---------	---	----	----	----	----	-------	-------	------	----	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	--

Rank Histogram



Relative Position (RELP) Histogram



Observation Error - Logic

- Account for observation error when verifying ensemble forecasts.
- Incorporate the method used by Meteo France into the Ensemble-Stat tool.
- May be applied to point observations or gridded analyses.
- Each observation value is compared to N ensemble member values.
- For each observation value, locate the observation error information (defined in config file or via a table lookup).
- Observation error information defines bias correction information and the assumed error distribution.
- ***For each observation value...***
 - Apply any bias correction logic.
- ***For each ensemble member value...***
 - Randomly sample from the error distribution centered on 0.
 - Add random perturbation to current ensemble member value.
 - Apply min/max criteria to keep perturbed value within valid physical limits.
- Perturbations are applied to the ensemble forecast values to make them comparable with the observations which already include errors.

Observation Error - MET Config

- Ensemble-Stat config file turns observation error perturbations ON/OFF.
- Applies the *same error distribution* to all observations for this task.

```
// Observation error options
// Set dist_type to NONE to use the observation error table instead
// May be set separately in each "obs.field" entry
obs_error = {
    flag                  = FALSE; // TRUE or FALSE
    dist_type             = NONE; // Distribution type (e.g. NORMAL)
    dist_parm             = []; // Distribution parameters (e.g. 2.0)
    inst_bias_scale       = 1.0; // Instrument bias scale adjustment
    inst_bias_offset      = 0.0; // Instrument bias offset adjustment
    min                  = NA; // Valid range of data
    max                  = NA;
}
```

- Bias Correction: $\text{obs_bc} = \text{obs} * \text{inst_bias_scale} + \text{inst_bias_offset};$

Observation Error - Table Lookup

- Observation error table lookup provides much finer control of the error distributions.
- In Ensemble-Stat config file, set “**flag = TRUE;**” and “**dist_type = NONE;**”
- If **MET_OBS_ERROR_TABLE** is set, read that table.
- If unset, read defaults from **MET_BASE/table_files/obs_error_table.txt**.
- Columns of observation error table file:
 - Filter criteria to find a match:
 - **OBS_VAR**: Comma-separated list of regular expressions for variable strings to match
 - **MESSAGE_TYPE**: Comma-separated list of message types to match
 - **PB_REPORT_TYPE, IN_REPORT_TYPE, INSTRUMENT_TYPE**: Additional filters
 - **STATION_ID**: Comma-separated list of Station ID's to match
 - **HGT_RANGE, PRS_RANGE, VAL_RANGE**: Numeric filter range. Set to **ALL** or min,max
 - Observation bias correction settings:
 - **INST_BIAS_SCALE**: Observation bias scale value (multiply)
 - **INST_BIAS_OFFSET**: Observation bias offset value (add)
 - Random perturbation settings:
 - **DIST_TYPE**: Distribution type (see User's Guide)
 - **DIST_PARM**: Parameters of distribution
 - **MIN, MAX**: Physical bounds for this variable

Observation Error - Example

- Default perturbations in `obs_error_table.txt` provided by Jeff Beck (NOAA/GSD) based on literature review.
- Defaults included for surface (**ADPSFC**) and upper-air (**ADPUPA**) message types for **TMP**, **DPT**, **WIND**, **UGRD**, and **VGRD**, as well as **HGT** and **APCP**.

```
APCP_[0-9]* ALL ALL ALL ALL ALL ALL ALL 0,0.01 NA NA NONE NA 0 NA
APCP_[0-9]* ALL ALL ALL ALL ALL ALL ALL 0.01,10 NA NA NORMAL 0.1 0 NA
APCP_[0-9]* ALL ALL ALL ALL ALL ALL ALL 10,50 NA NA NORMAL 1.0 0 NA
APCP_[0-9]* ALL ALL ALL ALL ALL ALL ALL 50,200 NA NA NORMAL 5.0 0 NA
APCP_[0-9]* ALL ALL ALL ALL ALL ALL ALL 200,500 NA NA NORMAL 15.0 0 NA
```

- **APCP_[0-9]*** uses regular expressions to match multiple accumulations.
- Lines match **ALL** message, pb report, input report, and instruments types.
- Lines match **ALL** station ID's and height and pressure ranges.
- Define different perturbations based on the value ranges, increasing from **NONE** for 0 - 0.1 mm up to **NORMAL(15)** for 200 - 500 mm.
- Min/Max values (**0 NA**) prevent negative perturbations.
- Increase Ensemble-Stat verbosity level (**-v 4**) to see perturbation details.

Observation Error - Output

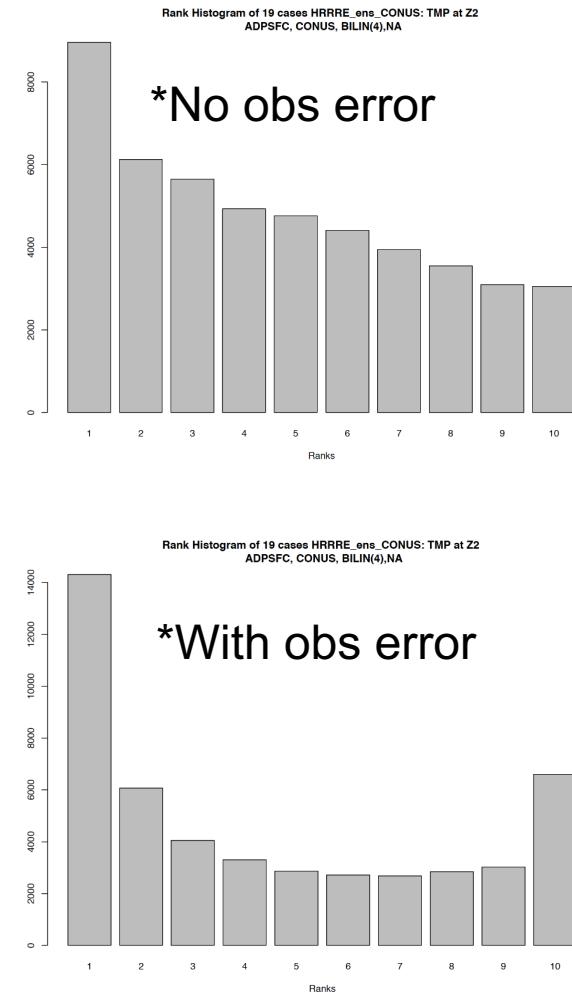
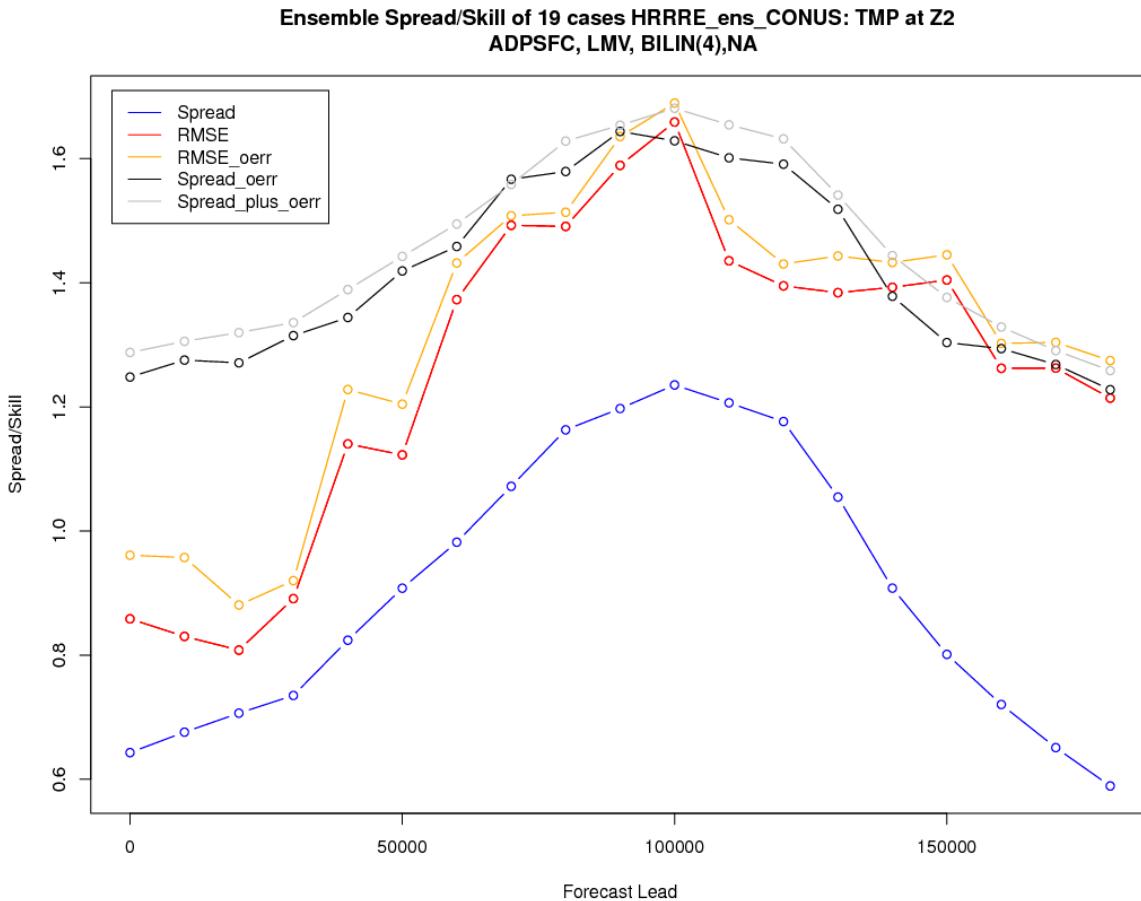
- Add new Ensemble Continuous Statistics (**ECNT**) line type.
- Move columns for CRPS, IGN, CRPSS, and SPREAD from **RHIST** to new **ECNT** line type.

Table 9.2: Format information for ECNT (Ensemble Continuous Statistics) output line type.

ECNT OUTPUT FORMAT		
Column Number	ECNT Column Name	Description
22	ECNT	Ensemble Continuous Statistics line type
23	TOTAL	Count of observations
24	N_ENS	Number of ensemble values
25	CRPS	The Continuous Ranked Probability Score
26	CRPSS	The ContinuousRanked Probability Skill Score
27	IGN	The Ignorance Score
28	ME	The Mean Error of the ensemble mean (unperturbed or supplied)
29	RMSE	The Root Mean Square Error of the ensemble mean (unperturbed or supplied)
30	SPREAD	The mean of the spread (standard deviation) of the unperturbed ensemble member values at each observation location
31	ME_OERR	The Mean Error of the PERTURBED ensemble mean (e.g. with Observation Error)
32	RMSE_OERR	The Root Mean Square Error of the PERTURBED ensemble mean (e.g. with Observation Error)
33	SPREAD_OERR	The mean of the spread (standard deviation) of the PERTURBED ensemble member values (e.g. with Observation Error) at each observation location
34	SPREAD_PLUS_OERR	The square root of the sum of unperturbed ensemble variance and the observation error variance

Observation Error - Plots

- When **obs_error** is enabled, perturbed data is used for all output line types:
 - ECNT, RHIST, PHIST, ORANK, SSVAR, RELP



* Images courtesy of Jeff Beck (NOAA/GSD and DTC)

Ensemble-Stat: Output Files

- netCDF
 - Gridded ensemble mean, standard deviation, min, max, range, frequency
 - “orank” file (gridded obs rank)
- ASCII
 - Up to 5 txt files and stat file
 - Continuous statistics (CPSS, IGN)
 - Ranked histogram
 - Probability integral transform histogram
 - Skill/spread variance
 - e.g. FBAR, OBAR, MSE, RMSE, PR_CORR
 - Relative position

```
ensemble_flag = {  
    mean      = TRUE;  
    stdev     = TRUE;  
    minus     = TRUE;  
    plus      = TRUE;  
    min       = TRUE;  
    max       = TRUE;  
    range     = TRUE;  
    vld_count = TRUE;  
    frequency = TRUE;  
    rank      = TRUE;  
    weight    = FALSE;  
};
```

```
output_flag = {  
    ecnt   = NONE;  
    rhist  = NONE;  
    phist  = BOTH;  
    orank  = BOTH;  
    ssvar  = BOTH;  
    relp   = BOTH;  
};
```

Ensemble Continuous Stats

Table 9.2: Format information for ECNT (Ensemble Continuous Statistics) output line type.

ECNT OUTPUT FORMAT		
Column Number	ECNT Column Name	Description
24	ECNT	Ensemble Continuous Statistics line type
25	TOTAL	Count of observations
26	N_ENS	Number of ensemble values
27	CRPS	The Continuous Ranked Probability Score
28	CRPSS	The ContinuousRanked Probability Skill Score
29	IGN	The Ignorance Score
30	ME	The Mean Error of the ensemble mean (unperturbed or supplied)
31	RMSE	The Root Mean Square Error of the ensemble mean (unperturbed or supplied)
32	SPREAD	The mean of the spread (standard deviation) of the unperturbed ensemble member values at each observation location
33	ME_OERR	The Mean Error of the PERTURBED ensemble mean (e.g. with Observation Error)
34	RMSE_OERR	The Root Mean Square Error of the PERTURBED ensemble mean (e.g. with Observation Error)
35	SPREAD_OERR	The mean of the spread (standard deviation) of the PERTURBED ensemble member values (e.g. with Observation Error) at each observation location
36	SPREAD_PLUS_OERR	The square root of the sum of unperturbed ensemble variance and the observation error variance

Ensemble Histograms

Table 9.3: Format information for RHIST (Ranked Histogram) output line type.

RHIST OUTPUT FORMAT		
Column Number	RHIST Column Name	Description
22	RHIST	Ranked Histogram line type
23	TOTAL	Count of observations
24	N_RANK	Number of possible ranks for observation
25	RANK_i	Count of observations with the ith rank (repeated)

Table 9.4: Format information for PHIST (Probability Integral Transform Histogram) output line type.

PHIST OUTPUT FORMAT		
Column Number	PHIST Column Name	Description
24	PHIST	Probability Integral Transform line type
25	TOTAL	Count of observations
26	BIN_SIZE	Probability interval width
27	N_BIN	Total number of probability intervals
28	BIN_i	Count of observations in the ith probability bin (repeated)

Table 9.5: Format information for RELP (Relative Position) output line type.

RELP OUTPUT FORMAT		
Column Number	RELP Column Name	Description
24	RELP	Relative Position line type
25	TOTAL	Count of observations
26	N_ENS	Number of ensemble members
27	RELP_i	Number of times the i-th ensemble member's value was closest to the observation (repeated). When n members tie, 1/n is assigned to each member.

Ensemble Obs Rank (matched pair)

Table 9.6: Format information for ORANK (Observation Rank) output line type.

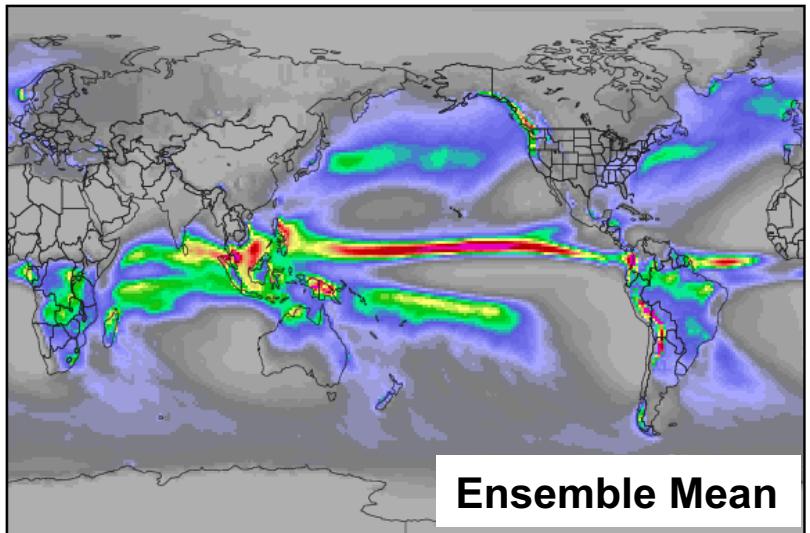
ORANK OUTPUT FORMAT		
Column Number	ORANK Column Name	Description
24	ORANK	Observation Rank line type
25	TOTAL	Count of observations
26	INDEX	Line number in ORANK file
27	OBS_SID	Station Identifier
28	OBS_LAT	Latitude of the observation
29	OBS_LON	Longitude of the observation
30	OBS_LVL	Level of the observation
31	OBS_ELV	Elevation of the observation
32	OBS	Value of the observation
33	PIT	Probability Integral Transform
34	RANK	Rank of the observation
35	N_ENS_VLD	Number of valid ensemble values
36	N_ENS	Number of ensemble values
37	ENS_i	Value of the ith ensemble member (repeated)
Last-6	OBS_QC	Quality control string for the observation
Last-5	ENS_MEAN	The unperturbed ensemble mean value
Last-4	CLIMO	The value of the included climatology
Last-3	SPREAD	The spread (standard deviation) of the unperturbed ensemble member values
Last-2	ENS_MEAN_OERR	The PERTURBED ensemble mean (e.g. with Observation Error).
Last-1	SPREAD_OERR	The spread (standard deviation) of the PERTURBED ensemble member values (e.g. with Observation Error).
Last	SPREAD_PLUS_OERR	The square root of the sum of the unperturbed ensemble variance and the observation error variance.

Binned Spread- Skill using Variance

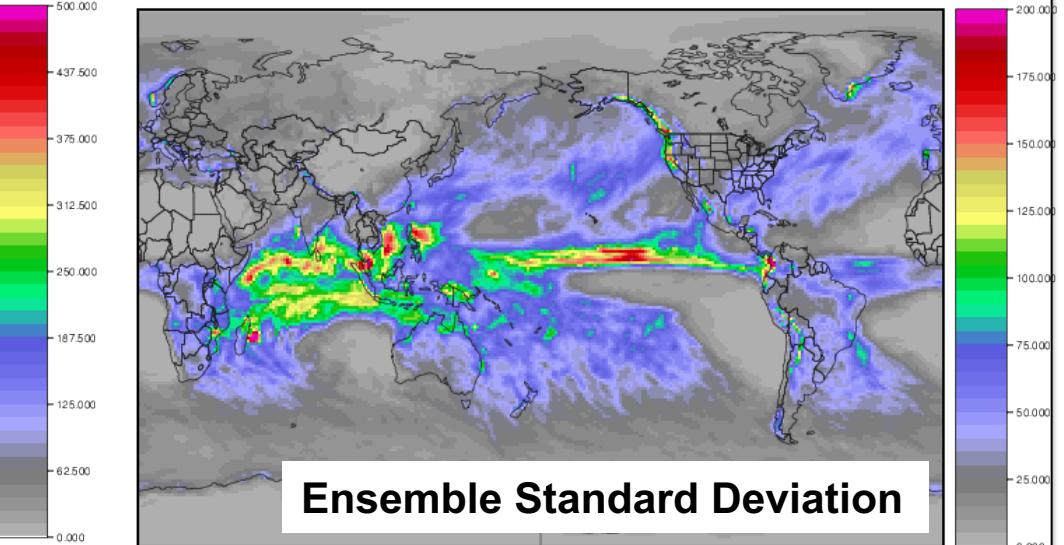
Table 9.7: Format information for SSVAR (Spread/Skill Variance) output line type.

SSVAR OUTPUT FORMAT		
Column Number	SSVAR Column Name	Description
24	SSVAR	Spread/Skill Variance line type
25	TOTAL	Count of observations
26	N_BIN	Number of bins for current forecast run
27	BIN_i	Index of the current bin
28	BIN_N	Number of points in bin i
29	VAR_MIN	Minimum variance
30	VAR_MAX	Maximum variance
31	VAR_MEAN	Average variance
32	FBAR	Average forecast value
33	OBAR	Average observed value
34	FOBAR	Average product of forecast and observation
35	FFBAR	Average of forecast squared
36	OOBAR	Average of observation squared
37-38	FBAR_NCL, FBAR_NCU	Mean forecast normal upper and lower confidence limits
39-41	FSTDEV, FSTDEV_NCL, FSTDEV_NCU	Standard deviation of the error including normal upper and lower confidence limits
42-43	OBAR_NCL, OBAR_NCU	Mean observation normal upper and lower confidence limits
44-46	OSTDEV, OSTDEV_NCL, OSTDEV_NCU	Standard deviation of the error including normal upper and lower confidence limits
47-49	PR_CORR, PR_CORR_NCL, PR_CORR_NCU	Pearson correlation coefficient including normal upper and lower confidence limits
50-52	ME, ME_NCL, ME_NCU	Mean error including normal upper and lower confidence limits
53-55	ESTDEV, ESTDEV_NCL, ESTDEV_NCU	Standard deviation of the error including normal upper and lower confidence limits
56	MBIAS	Magnitude bias
57	MSE	Mean squared error
58	BCMSE	Bias corrected root mean squared error
59	RMSE	Root mean squared error

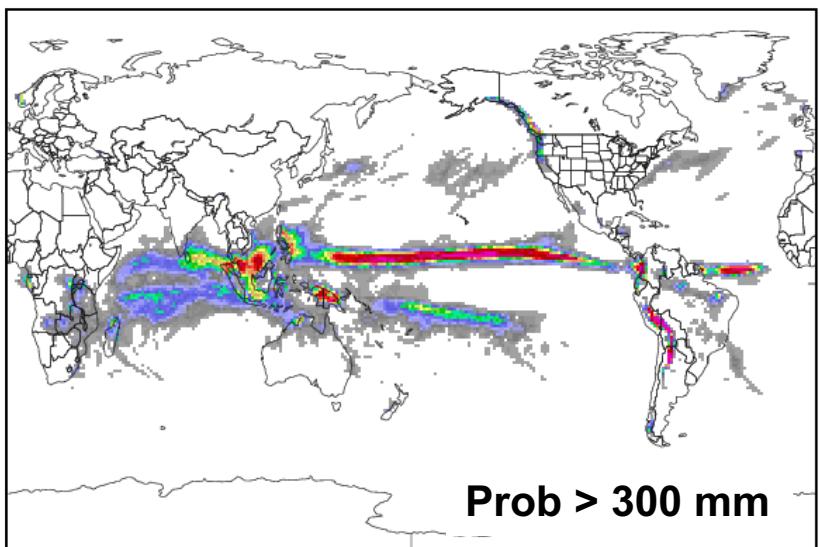
Ensemble Stat Tool: nc Output



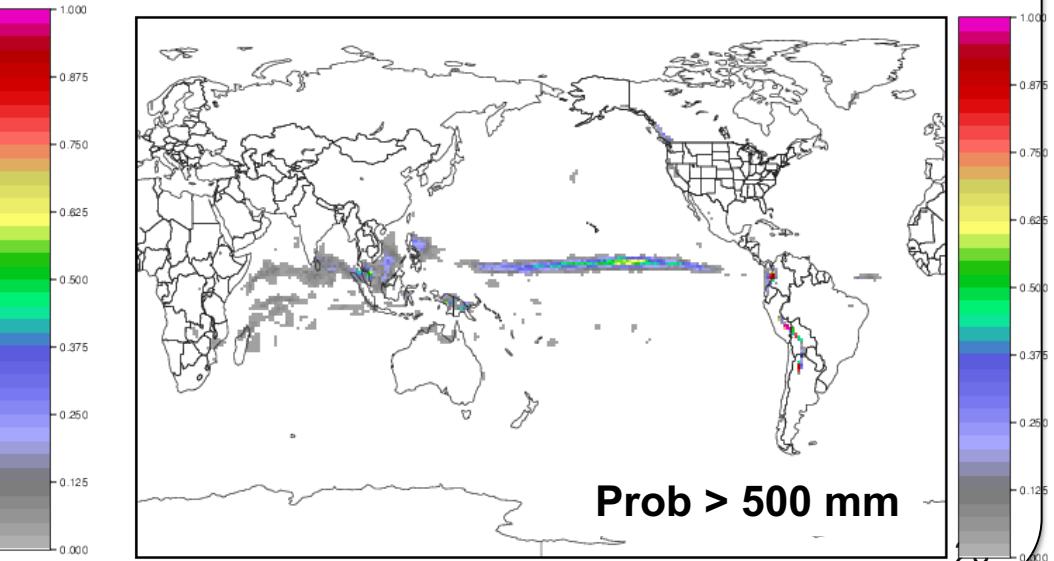
Ensemble Mean



Ensemble Standard Deviation



Prob > 300 mm



Prob > 500 mm

Uses for Output from Ensemble Stat

