



**WESTERN REGION TECHNICAL ATTACHMENT  
NO. 97-38  
DECEMBER 2, 1997**

---

**THE "EVERIFY" TEMPERATURE AND PRECIPITATION  
FORECAST VERIFICATION PROGRAM**

**David R. Eversole - NWSO Mobile, Alabama**  
(Reprinted from Southern Region Technical Attachment)

**Introduction**

With the integration of UNIX-based workstations into the forecasting environment of the National Weather Service (NWS), the author recognized the need for an automated forecast verification program for temperature and precipitation forecasts which would run on a UNIX platform. The resulting program is entitled *Everify* and actually consists of two programs (*Everify* and *Everify\_capture*), and a configuration file (*Everify\_config*). These programs, specifically the *Everify\_capture* program, are completely automated and require little maintenance. In addition, these programs initiate and maintain a database of human forecasts, various model output statistics (MOS), and verifying data that can be easily manipulated at any time.

***Everify***

One of the main reasons for writing the *Everify* programs was to bridge the time gap for spin up offices between Automation of Field Operations and Services (AFOS) and Advanced Weather Interactive Processing System (AWIPS) forecaster verification programs. At this time, the AWIPS forecaster verification program is still being developed, so *Everify* may well be useful for some time to come even after the delivery of AWIPS. Of the two programs contained in *Everify*, the *Everify* program itself is designed to perform "on demand" verification and statistical information concerning forecast temperatures and precipitation probabilities (POP's), and also has the ability to perform ranking and analysis of forecaster/MOS performance and biases, all in an easy to understand format, as well as several other features.

The second component of *Everify* is the *Everify\_capture* program which performs decoding of METAR observations, FWC (NGM MOS output) and FAN (AVN MOS output) guidance products, and either the SFD (State Forecast Discussion), AFD (Area Forecast

Discussion) or CCF (Coded Cities Forecast) product. The program strips the synoptic temperature and precipitation information from the METAR observations and produces highs, lows and precipitation amounts in accordance with national/regional verification standards. For example, Southern Region Headquarters (SRH) ROML S-4-84 states that highs and daytime precipitation are from 1200 UTC to 0000 UTC, nighttime precipitation is from 0000 UTC to 1200 UTC and lows are from 0000 UTC to 1800 UTC. In the event of a cold frontal passage, an editor can be used to adjust the temperature to that which is closest to 8 am local time, if desired. The *Everify\_capture* program also performs numerous quality control checks during processing, and will send a selected individual an electronic mail message concerning problems encountered during decoding and processing. One quality control feature of *Everify\_capture* is that the existence of current or past precipitation is checked in each hourly observation of a synoptic period (0000 UTC to 0600 UTC for example) when a hundredth of precipitation is carried in the synoptic observation for that period (0600 UTC is this example). If a discrepancy is found, then *Everify\_capture* records zero precipitation for that period, if directed to do so in the *Everify\_config* configuration file. This quality control check is recommended due to the fact that Automated Surface Observing System (ASOS) installations occasionally report an erroneous hundredth of precipitation, usually during dense fog events and/or windy events, but may also be due to other factors.

In order for the *Everify* programs to work, an LDM software link to the AFOS data feed must exist (The LDM software is free; see your ESA or contact the Unidata Program Center which is managed by UCAR in Boulder, CO for more information). Once the LDM link exists, the *pqact* file in the *~/etc* directory under the LDM home directory will require a simple modification to send the guidance and METAR observations for each verifying station as well as the AFD, SFD or CCF.

The *Everify* programs will benefit any office which desires to initiate automated forecast verification, and has the following advantages:

- easy to understand descriptive statistics
- runs in a UNIX environment
- is easily configured for verification of up to 9 stations
- can decode the following products: AFD, SFD or CCF for use in verification
- allows for "ghost" guidance stations. This would be for the case where the guidance output (MOS) is for a station near the verifying station, but not for the verifying station itself. When instructed to do so in the *Everify\_config* configuration file, the *Everify\_capture* program will then treat the guidance data as if it were for the verifying station itself, thus "ghosting" the guidance data for the verifying station. The *Everify\_config* file will allow for up to 9 such stations.

- allows for "ghost" verifying stations. This is the case in which the verifying station does not provide synoptic data, but a nearby station does. When instructed to do so in the *Everify\_config* configuration file, the *Everify\_capture* program will then treat the synoptic data as if it were for the verifying station, thus "ghosting" the synoptic data for the verifying station. The *Everify\_config* file will also allow for up to 9 such stations.

- a password protected editor is provided which can be used to correct and/or enter observed, guidance and/or forecast data. The password also allows for some degree of security in order to protect the authenticity of the guidance, forecast and verifying data.

- a password protected forecaster performance analysis is available which produces a descriptive statistical analysis of forecaster temperature and POP biases, as well as rankings among other forecasters and guidance (FWC, FAN) based upon forecaster skill. The analyses are very versatile and can be produced for any year or all years, any range of 12 or less months, day shift issued forecasts, night shift issued forecasts or both, and any individual forecast station, or all stations. As a result, the performance analyses provide highly detailed information concerning forecaster/guidance biases and skill for: 1) location, 2) each of the first three periods, and 3) cold/warm/transitional seasons.

- a highly detailed individual forecaster performance analysis is also available. This analysis is for a single forecaster, and is hence not password protected as no comparisons are made between forecasters. The analysis can be produced for any year or all years, any range of 12 or less months, day shift issued forecasts, night shift issued forecasts or both, and any individual forecast station, or all stations. Histograms are provided for all POP's, POP's where at least a hundredth of precipitation occurred, and POP's where a trace or less of precipitation occurred. These histograms will enable the analysis of climate, as well as forecaster POP biases, and forecaster preferences among POP categories (0,5,10,20, etc). A POP table is also provided which shows the percentage of precipitation forecasts verified for each POP category, as well as the average POP when at least a hundredth of precipitation occurred, the average POP when a trace or less of precipitation occurred, and a wet/dry bias indicator. A histogram is also produced for degrees of temperature forecast error. All of the above tables are produced for each of the first three periods.

- a decoder is available which will decode bulk observations downloaded from an ASOS site. This decoder will quickly process the data, check for any erroneous hundredths of precipitation, check for corrected and/or duplicate observations, and produce a table of daytime highs and precipitation as well as nighttime lows and precipitation. A listing is also provided of each of the observations which contained erroneous hundredths of precipitation.

## Everify Output Statistics

The Everify programs produce descriptive statistics, not inferential statistics. Before showing examples of the statistical output, the following is a brief description and review of the statistical equations used to produce the statistics.

### - Mean Algebraic Error:

$$\frac{1}{n}\sum(F-O)$$

- The mean algebraic (or average as it is commonly known) error is used to describe a warm or cool temperature bias. It is also used in the forecast skill analysis to describe a dry or wet POP bias.

### - Mean Absolute Error:

$$\frac{1}{n}\sum|F-O|$$

- The mean absolute error is used to describe the magnitude of temperature error. It is not as descriptive as the root mean squared error (shown below) due to the fact that the summed errors are not squared. It is nevertheless, a useful statistic.

### - Root Mean Square Error:

$$\sqrt{\frac{1}{n}\sum(F-O)^2}$$

- The Root Mean Square (RMS) error is used as a measure of error in temperature forecasts. Since each error from the observed value is squared, RMS amplifies the larger errors. This amplification of larger errors creates a larger separation between forecasters with differing levels of skill, and is thus a better descriptor for temperature error.

### - Brier Score:

$$\frac{1}{n}\sum(F-O)^2$$

- The Brier Score is very similar to the root mean square in that the precipitation probability error is squared. The Brier Score is a good statistic for describing the precipitation probability error since the error is squared, thus amplifying large errors and widening the separation between forecasters of differing levels of skill. The variable "F" is the forecasted probability of precipitation (POP), and has values between 0 and 1. The variable "O" indicates whether or not at least a hundredth of rainfall occurred, and has a value of 0, for no precipitation, or 1, for a hundredth or more of precipitation. NWS forecasters

actually issue POP's in the form of chances, which have values between 0 and 100%, so to use this equation using chances, the variable "O" will have a value of either 0 or 100.

- Brier Skill Score:

$$\frac{B_{MOS} + B_{Forecaster}}{B_{MOS}} * 100$$

- The Brier Skill Score is a measure of the forecaster's percent improvement over MOS (either FWC or FAN). Positive values indicate a superior performance compared to MOS, while negative values indicate a degraded performance compared to MOS.

- Standardized Random Variables:

$$\frac{X - \mu}{\sigma}$$

- In the analysis of forecaster skill, the RMS temperature error and the Brier scores are standardized. By standardizing both of these, the mean of each set of numbers becomes zero (0), and the standard deviation becomes one (1). The standardized temperature RMS error and standardized Brier scores can then be averaged (using a weighting function based upon the number of valid data for each), to produce a third set of numbers which indicates the forecasters combined temperature and precipitation forecast skill.

The following are examples of statistical output from the *Everify* program. For brevity, only a portion of the monthly data is shown in the examples. Note that this information can be produced for any or all forecasters, for the current month, or any past month in which the *Everify* program has created data. Table 1 is an example of the monthly summary product. This gives a tabulation of the observed high and low temperatures for each day as well as observed precipitation data. It is this data which is used to verify the forecast and guidance data.

**Table 1. Example of the Monthly Summary Table**

--- Summary of Temperature and Precipitation Data for MOB on 03/1997 ---

-DATE-	---- Daytime ----		--- Overnight ---	
	High	Precip	Low	Precip
03/01/97	81	0	70	0
03/02/97	77	0	59	T
03/03/97	76	0	52	0.01

Table 2 is an example of the "Forecast" versus "Observed" output. This gives a quick look at the performance of the forecaster as compared to the observed data. This table is actually one of two produced, and is for the day shift (1200 UTC MOS guidance) temperature forecasts, where the first valid period is for that night. The other table that would be produced, but is not shown, would be for night shift (0000 UTC MOS guidance) temperature forecasts. This format of dividing the tables between the two forecast periods was done in order to match some of the older manual forecast verification techniques and also to aid in the interpretation of the data. This table can be produced for either all forecasts, as is shown, or for any one particular forecaster.

The mean temperature error is provided to indicate biases in the forecaster's temperature forecasts, while the mean absolute temperature error indicates the amount of the forecast error in temperature. In the table, "fn" is the forecaster number, "ft" is the forecasted temperature, "at" is the actual (observed) temperature, "f-o" is the difference between the forecasted and actual (observed) temperature, "POP" is the forecasters POP, and "R?" is whether or not a hundredth inch or more rain fell. Note that if some of the verifying (observed) data are missing since this data has not yet been observed, then the observed data will be marked as missing ("MM" or "M" as appropriate) until the verifying data comes in. Periods with missing data are not considered in either the mean or mean absolute temperature error statistics (which are included at the base of the columns).

Tables 3 and 4 are examples of two of four Forecast/Guidance vs. Observed tables. Table 3 is for the temperature forecasts made on day shifts (first verifying period being that night), while the last table is for precipitation (POP) forecasts made on day shifts. The other two tables that are produced, but not shown, are the temperature and precipitation forecasts made on the night shift (first verifying period being that day). The *Everify* program can be easily run to display data for all forecasts, such as in this example, or for any one particular forecaster.

**Table 2. Example of the Forecast vs. Observed Table**

--- Verification of Daytime Forecasts for MOB on 03/1997 ---

-DATE-	----- Tonight -----					----- Tomorrow -----					----- Tomorrow Night -----				
	fn	ft	at	f-o	POP R?	ft	at	f-o	POP R?	ft	at	f-o	POP R?		
03/01/97	40	69	70	-1	40% N	74	77	-3	80% N	58	59	-1	30% N		
03/02/97	40	59	59	0	70% N	71	76	-5	0% N	54	52	2	0% Y		
03/03/97	40	52	52	0	0% Y	76	78	-2	0% N	62	66	-4	0% N		
Mean t err			-0.3					-3.3					-0.3		
Mean Abs. Temp err			0.3					3.3					2.3		

**Table 3. Example of the Forecast/Guidance Temperatures vs. Observed Temperatures Table**

--- Verification of Day Shift Temperature Forecasts for MOB on 03/1997 ---

day	----- Tonight -----						---- Tomorrow ----					-- Tomorrow Night --				
	fn	fwc	fan	ft	ot	f-o	fwc	fan	ft	ot	f-o	fwc	fan	ft	ot	f-o
01	40	67	65	69	70	-1	75	72	74	77	-3	56	54	58	59	-1
02	40	61	62	59	59	0	76	75	71	76	-5	52	53	54	52	2
03	40	55	49	52	52	0	77	74	76	78	-2	65	64	62	66	-4
mn err		0.7	-1.7	-0.3			-1.0	-3.3	-3.3			-1.3	-2.0	-1.0		
abs er		2.7	3.7	0.3			1.0	3.3	3.3			1.3	2.7	2.3		
rms er		2.7	3.8	0.6			1.3	3.7	3.6			1.8	3.2	2.6		

  

	---Tonight---	---Tomorrow---	---Tomorrow Night---
Improvement over FWC	2.1 deg / 78%	-2.3 deg / -176%	-0.8 deg / -44%
Improvement over FAN	3.2 deg / 84%	0.1 deg / 3%	0.6deg/ 19%

Note: Negative values represent a lower performance than FWC/FAN

The key to Table 3 is as follows: "fn" is the forecaster number, "fwc" and "fan" are the guidance data, "ft" is the forecaster's temperature, "ot" is the observed (actual) temperature, and "f-o" is the difference between the forecasted and observed temperature. For the statistics provided at the bottom of the table, "mn err" is the mean temperature error, "abs er" is the absolute temperature error, and "rms er" is the root mean square temperature error. Since the root mean square error is a better descriptor of temperature error (temperature differences are squared in the summation process), this statistic is used in the "Improvement over FWC" and "Improvement over FAN" summary statistics at the bottom of the table. Note that if not all of the verifying (observed) data have yet come in, then those periods will denoted with "MM." Periods with missing data are not considered in the statistical calculations at the base of the columns.

**Table 4. Example of the Forecast/Guidance POP's vs. Observed Precipitation Table**

--- Verification of Day Shift Precipitation Forecasts for MOB on 03/1997 ---

Day	----- Tonight -----				---- Tomorrow ----				-- Tomorrow Night --			
	fn	fwc	fan	pop rain	fn	fwc	fan	pop rain	fn	fwc	fan	pop rain
01	40	33	38	40 0	71	82	80	0 0	26	49	30	0 0
02	40	66	88	70 0	34	22	0	0 0	54	55	0	0.01
03	40	53	12	0 0.01	40	29	0	0 0	40	76	0	0
brier	2551	5644	5500		2599	2683	440		1464	3400	3633	

  

	Tonight	Tomorrow	Tomorrow Night
FWC Brier Skill Score	-115.6%	83.1%	-148.2%
FAN Brier Skill Score	2.6%	83.6%	-6.8%

Note that negative values represent a lower performance than FWC/FAN.

Note that the root mean square error (as well as mean error and absolute error) are for **that column only**. Manual calculations of the improvement over FAN or FWC may or may not equal the percentages given in the improvement section at the bottom of the table. This is due to the fact that the mean, absolute and root mean square errors are computed for each of the columns, independent of the others. For example, for the FWC guidance, the mean, absolute and root mean square errors are for the FWC guidance **only**. When the improvement over guidance values are computed, valid guidance, forecast and observed data must be present for each day in the period of consideration before that particular day's period data will be used. This is necessary in that if guidance data was missing during a particularly error ridden set of forecasts, but was present and accurate for the other days, then the improvement score would inaccurately show that the forecaster had a lower performance than guidance. In this manner, the improvement statistics represents the most accurate measure of forecaster performance versus the guidance products.

The key to Table 4 is as follows: "fn" is the forecaster number, "fwc" and "fan" are the guidance data, "pop" is the forecaster's POP, and "rain" is the actual amount of rainfall. The Brier Score statistic is located at the bottom of the columns, entitled "brier." As in Table 3, when the Brier Skill Scores are computed ("FWC Brier Skill Score" and "FAN Brier Skill Score"), valid guidance, forecast and observed data must be present for each day in consideration before that particular day's data will be used. As in the Table 3, this allows for the improvement statistics to represent the most accurate measure of forecaster performance versus the guidance products.

Tables 5 is an example of the Temperature Forecast Skill Ranking table that was generated for day shift forecasts (afternoon package) for the months from January to April at station MOB (Mobile, AL). When the forecaster ranking analysis is run using *Everify*, three tables are created: The Temperature Forecast Skill Ranking Table (example: Table 5), the Precipitation Forecast Skill Ranking Table (not shown), and the Combined Temperature and Precipitation Forecast Skill

### **Example of the Temperature Forecast Skill Ranking Table**

```
>>> This forecaster ranking was run on 04/30/97 at 06:33Z <<<
>>> For Day shift issued AFD/SFD/CCF's temperature forecasts <<<
>>> for 1997 with months JAN-APR at all stations <<<
```



>>> First period (Tonight)...

Fcstr	RMS Error	Standardized RMS	Bias (Avg)	# fcsts
40	2.56	-1.33	-0.33	9
41	3.13	-0.70	-0.20	15
FAN	3.71	-0.05	0.36	250
FWC	3.80	0.05	0.44	392
42	6.14	2.67	1.50	10

>>> Second period (Tomorrow)...

Fcstr	RMS Error	Standardized RMS	Bias (Avg)	# fcsts
42	2.92	-1.57	0.60	86
41	3.36	-0.96	-0.03	90
FWC	4.12	0.08	0.20	388
FAN	4.13	0.10	-0.08	246
40	4.46	0.55	-0.78	87

>>> Third period (Tomorrow Night)...

Fcstr	RMS Error	Standardized RMS	Bias (Avg)	# fcsts
41	3.28	-1.59	0.37	94
FWC	4.53	-0.36	-0.10	382
FAN	4.72	-0.17	-0.38	244
40	5.10	0.20	-0.21	119
42	5.17	0.27	-1.92	13

>>> And for all three periods combined...

Fcstr	RMS Error	Standardized RMS	Bias (Avg)	# fcsts
41	3.52	-1.05	-0.23	270
42	3.76	-0.69	0.85	333
FWC	4.15	-0.11	0.18	1162
FAN	4.18	-0.06	-0.03	740
40	4.23	0.51	-0.37	27

Ranking Table (not shown). The ranking analysis displays not only the level of skill for each forecaster (including FWC and FAN guidance), but also displays temperature and precipitation (including wet, dry and both) forecast biases for each period and all periods combined. By including the ranking of guidance, easy comparisons are made as to which forecasters possess a skill greater than guidance. The ranking analysis can be run for any or all years, any month or period of months, either or both forecast package issuances

(early morning or afternoon package), and for any or all forecast stations. This flexibility allows for detailed examination of forecaster strengths and weaknesses.

The key to Table 5 is as follows: "Fcstr" is the forecaster number, "RMS Error" is the root mean square temperature error, ranked in order of superior skill (lowest RMS), "Standardized RMS" is the standardized root mean square error, "Bias (Avg)" is the mean temperature error, and "# fcsts" is the number of forecasts *with verifying data* for the period(s) of interest. The standardized RMS is used to produce the combined temperature/precipitation table, and the Bias is used to indicate either a warm (positive values) or cool (negative values) temperature bias.

## Future plans

Additional improvements that will be coming soon to the *Everify* programs:

- 1) Adjustable guidance performance statistics - for example, an indication of biases in guidance over a selectable number of days.
- 2) Flexibility in the period of determination of highs, lows and precipitation - this will allow for the *Everify* programs to be used in the western portion of the United States, and possibly the Alaska region.
- 3) Conversion of the c-shell code in which the programs are written to C++. This will improve the operating speed of the program.
- 4) Verification of winds and seas in Coastal Marine Forecasts (CWF's).
- 5) Verification of (Terminal) Aerodrome Forecasts (TAF's).

## References

- Hughes, Lawrence, A., 1980: National Weather Service, Probability Forecasting - Reasons, Procedures, Problems. NOAA Technical Memorandum NWS FCST-24.
- Jensen, Ray, 1984: National Weather Service, Public Forecast Verification. Southern Region Operations Manual Letter S-4-84.
- Unidata, 1995: Unidata LDM5 - Site Managers Guide, Unidata Program Center, Boulder, CO, UCAR.
- Winkler, Robert, L., and Hays, William, L., *Statistics - probability, inference, and decision*. 2d Ed. New York: Holt, Rinehart and Winston, Inc., 1975.