NATIONAL WEATHER SERVICE INSTRUCTION 10-1601
February 24, 2009

Operations and Services

Performance, NWSPD 10-16

VERIFICATION PROCEDURES

NOTICE: This publication is available at: http://www.nws.noaa.gov/directives/.

OPR: W/OS52 (C. Kluepfel) **Certified by:** W/OS5 (C. Woods)

Type of Issuance: Routine

SUMMARY OF REVISIONS: This directive supersedes National Weather Service Instruction 10-1601, dated November 3, 2005. The following changes have been made to this directive:

1) The material describing the legacy (AFOS-Era) public forecast and TAF verification programs have been removed from sections 1 and 6, respectively. The material on *Aviation Verify* has been removed from section 6. These programs are no longer operational.

2) The valid periods for Probability of Precipitation forecasts have been changed (section 1.1.5).

3) Sections 1.4.1 and 1.4.2 (from Red Flag Warnings) were rewritten.

4) The verification of winter weather warnings (section 1.5), high wind warnings (section 1.6), and coastal/lakeshore flood warnings (section 3.2) has been automated; the statistics are now available from *Stats on Demand.* Event specific verification statistics (e.g., heavy snow, ice storm) are available for winter weather warnings. Regional reporting requirements of verification statistics for winter weather warnings, high wind warnings, and coastal flood/lakeshore warnings have been discontinued; this information is now available through *Stats on Demand.*

5) Storm-based verification is now available for severe thunderstorm and tornado warnings (section 2.2), special marine warnings (section 3.4), and flash flood warnings (section 4.2).

6) The time of flash flood warning issuance is taken from the Valid Time and Event Code (VTEC) line (section 4.1.3).

7) TAF verification TEMPO reports have been enhanced and clarified (section 6.1.6.3).

8) The section on the equitable skill score was updated (appendix A, section 2.9).

/signed/ February 10, 2009

David B. Caldwell Date
Director, Office of Climate,
Water, and Weather Services

Verification Procedures

Table of Contents: **Page**

Appendices

1. Public and Fire Weather Forecast Verification Procedures.

1.1 Public Forecast Verification at Points. The National Oceanic and Atmospheric Administration's (NOAA) National Weather Service (NWS) public forecasts issued by all Weather Forecast Offices (WFO) are verified at select points.

1.1.1 Verification Sites. All sites forecast in the point forecast matrices (PFM) that issue routine Meteorological Aviation Reports (METAR) and Special Aviation Weather Reports (SPECI) are verified unless the local WFO determines that a particular site is unrepresentative of its surroundings or inappropriate for verification. An interactive directory of all active verification sites is maintained on the NWS Performance Management website. The NWS seeks to incorporate all available observations into the verification program if the data meet NWS observation standards. See NWSI 10-1302, Instrument Requirements and Standards for the NWS Surface Observing Programs (Land).

1.1.2 Data Input. Public forecast data come from the scheduled PFMs issued by each WFO twice a day at 0400 and 1600 Local Time (LT). The latest 0400 (1600) LT PFM issued between 0000 and 0559 (1200 and 1759) LT, including corrections, are accepted. Amendments are not verified. Guidance forecasts come from the alphanumeric Model Output Statistics (MOS) guidance products derived from the following models: the North American Mesoscale (NAM), and the Global Forecast System (GFS). The verifying observations primarily come from all METAR/SPECI reports issued for each location in the PFMs. The satellite cloud product is also used as an observation source. All METARs and SPECIs are tested for reliability and consistency, and suspicious data are removed. These quality assurance algorithms are found on the NWS Performance Management website. Public forecast verification statistics are available back to January 2004.

Forecast data in the PFMs issued at 0400 LT, when matched to MOS guidance, are matched to the 0000 UTC cycle from the same date. Forecast data in the PFMs issued at 1600 LT, when matched to MOS guidance, are matched to the 1200 UTC cycle from the same date for 12-hour forecast periods 1 through 5. Beyond period 5, the forecast data in the PFMs issued at 1600 LT are matched to the 0000 UTC cycle from the same date.

1.1.3 Projections. Projections for public elements are defined in terms of the number of 12-hour forecast periods that have elapsed since the forecast issuance time (approximately 0400 and 1600 LT). Unless otherwise stated for the individual element, these 12-hour forecast periods are defined as 0600 to 1800 LT and 1800 to 0600 LT. Forecasts are made out to Day 7, totaling 13 or 14 projections in a single PFM.

1.1.4 Public Forecast Verification Reports. NWS employees access verification statistics from the *Stats on Demand* feature of the NWS Performance Management website. This Web page is operated and maintained by the Office of Climate Water and Weather Services (OCWWS) Performance Branch. *Stats on Demand* accesses an interactive database of monthly data and generates verification statistics customized to the user's request. The user requests data for any public weather element and, if desired, matching forecasts from a single MOS guidance product for one or more:

a. Months.

b. Forecast issuance times, i.e., early morning, late afternoon.

c. Forecast projections.

d. Verification sites, i.e., single site, multiple sites, regional, or national.

1.1.5 Elements.

a. Max/Min Temperatures. The forecast period for all daytime maximum temperatures is defined as 7 a.m. to 7 p.m. Local Standard Time (LST). The forecast period for all nighttime minimum temperatures is defined as 7 p.m. to 8 a.m. LST.

(1) Projections: Projections are expressed in 12- or 13-hour forecast periods, totaling 13 [14] in the PFM issued in the early morning [late afternoon]. All are verified.

(2) WFO Forecasts and MOS Guidance: Daytime maximum and nighttime minimum temperatures are forecast in whole degrees Fahrenheit out to Day 7.

(3) Observations: Daytime maximum and nighttime minimum temperatures are inferred from the METAR/SPECIs to the nearest degree Fahrenheit.

(4) Statistics: The following statistics are computed. As mentioned in section 1.1.2, the user has the option to select one guidance product for matching with the PFM, and when selected, this guidance product is always matched to the PFM at the appropriate time and place.

- Number of cases in the sample.

.

- Mean absolute error (MAE) for the entire sample (defined in appendix A, section 4.1); the percent improvement of the PFM over the selected guidance product is also provided for this statistic.

- Mean (algebraic) error (ME), see appendix A, section 4.1.

- Root mean square error (RMSE), see appendix A, section 4.1.

- Number of cases when the absolute error of the PFM or guidance was greater than or equal to 6º F.

- MAE when the absolute error of the PFM or guidance was greater than or equal to 6º F.

- Number of cases when the PFM, the selected guidance product, or the observation changed (in either direction) from the previous 24 hours by at least 10º F. Data are not provided for the first two 12-hour forecast projections.

- MAE when the PFM, the selected guidance product, or the observation changed (in either direction) from the previous 24 hours by at least 10º F; the percent improvement of the PFM over the selected guidance product under these circumstances is also provided. Data are not provided for the first two 12-hour forecast projections.

- Number of cases when the PFM was changed from the selected guidance product by 4º F or greater.

- MAE when the PFM was changed from the selected guidance product by 4º F or greater; the percent improvement of the PFM over the selected guidance product under these circumstances is also provided.

- For minimum temperatures only, when the previous day's minimum temperature was 40º F or greater, the following statistics are provided for forecast temperatures equal to or less than 32º F: probability of detection (POD), false alarm ratio (FAR), and critical success index (CSI). See Appendix A, section 3, for the definitions of POD, FAR, and CSI.

- Histogram of the absolute errors for PFM and the selected guidance product using the following absolute error categories in degrees Fahrenheit: 0-3, 4-5, 6-10, 11-15, greater than 5, greater than 10, and greater than 15. The value of each error category is provided as a percentage of the total sample. The percent improvement of the PFM over guidance is provided for the following absolute error categories: greater than 5º F and greater than 10º F.

b. Probability of Precipitation (PoP). Probability of 0.01 inch or greater liquid equivalent precipitation within the following 12-hour periods: 0000 to 1200 Universal Coordinated Time (UTC) and 1200 to 0000 UTC, except for Pacific Region where the valid periods at 0600 to 1800 UTC and 1800 to 0600 UTC.

(1) Projections: Projections are expressed in terms of 12-hour forecast periods, totaling 13 [14] in the PFM issued in the early morning [late afternoon]. All are verified.

(2) WFO Forecasts and MOS Guidance: The following forecast values are allowed in the PFM: {0, 5, 10, 20, 30, ..., 80, 90, 100}. MOS PoPs, forecast to the nearest percent, are rounded to the nearest allowable value.

(3) Observations: From METARs, 12-hour precipitation amounts to the nearest hundredth of an inch are inferred for the aforementioned periods. All precipitation gage reports are automatically quality controlled using the following: (a) internal consistency checks with other parts of the METAR report, (b) Stage III quantitative precipitation estimates issued by the River Forecast Centers, and (c) data from the national snow analysis issued by the National Operational Hydrologic Remote Sensing Center.

(4) Statistics: The following statistics are computed for 12-hour PoP forecasts. As mentioned in section 1.1.2, the user has the option to select one guidance product for matching with the PFM, and when selected, this guidance product is always matched to the PFM at the appropriate time and place.

- Number of forecast periods.
- Number of observed precipitation cases.
- Observed precipitation frequency; (b) divided by (a).
- Mean PoP Forecast: the mean PoP value for all chosen forecasts.
- Mean PoP Forecast with Precipitation: the mean PoP value for all chosen forecasts whenever 0.01 inch or greater (measurable) precipitation occurred.
- Mean PoP Forecast without Precipitation: the mean PoP value for all chosen forecasts whenever no measurable precipitation occurred.

- Brier score (defined in Appendix A, section 4.2); the percent improvement of the PFM over the selected guidance product is also provided.
- Brier Score when PFM PoP was 30% or greater; the percent improvement of the PFM over the selected guidance product under these circumstances is also provided along with the number of cases in parentheses.
- Brier score whenever measurable precipitation occurred; the percent improvement of the PFM over the selected guidance product under these circumstances is also provided along with the number of cases in parentheses.
- Brier score when PFM differed from selected guidance product by 20% or more; the percent improvement of the PFM over the selected guidance product under these circumstances is also provided along with the number of cases in parentheses.
- PoPs are interpreted as binary (yes/no) forecasts for measurable precipitation. PoPs greater than or equal to 50% are interpreted as "yes." PoPs less than 50% are interpreted as "no."
- The relative frequency of measurable precipitation is provided for the times when the following PoP thresholds were forecast: 0, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, and 100 percent. Probabilistic forecasts are perfectly reliable when each of the PoP thresholds equals the relative frequency of measurable precipitation events that occurred when it was forecast. For example, if measurable precipitation occurs 30% of the time that you forecast a 30% PoP, your 30% PoP forecasts were reliable.

1.2 National Digital Forecast Database (NDFD) Verification. MDL verifies the NDFD out to Day 7.

a. The following methods are used:

(1) Grid-to-Point. Only forecasts at the grid point nearest a METAR site are verified.

(2) Grid-to-Grid. All grid points are verified from the 5-kilometer Real Time Mesoscale Analysis (RTMA). The verification data are experimental.

b. The following elements are verified out to 7 days:

(1) Max/Min Temperature. Forecast periods are defined in the same manner as other public verification, i.e., 7 p.m. to 8 a.m. LST for minimum temperature, 7 a.m. to 7 p.m. LST for maximum temperature.

(2) 12-hour PoP. Forecast periods are defined 0000-1200 and 1200-0000 UTC.

(3) Temperature. Every 3 hours out to 72 hours; then every 6 hours out to 7 days.

(4) Dew point. Every 3 hours out to 72 hours; then every 6 hours out to 7 days.

(5) Wind direction and speed. Every 3 hours out to 72 hours; then every 6 hours out to 7 days.

Data are updated monthly and may be found on a website operated by MDL.

1.3 National Fire Danger Rating System (NFDRS) Forecast Verification. Forecasts and observations in this automated program come from the fire weather product with the AWIPS product identifier (PIL) NMCFWOrr, NMCFWOss or NMCFWOxxx, where rr refers to one of the four CONUS NWS regions, ss refers to a state, and xxx refers to a specific WFO.

Both NFDRS forecasts and observations are valid at 1300 LST but are issued as separate bulletins with the same product name. The forecasts are issued approximately 22 hours prior to the forecast valid time, and the verifying observations are disseminated shortly after 1300 LST the next day. For example, a forecast valid at 1300 LT will be issued at approximately 1500 LT the previous day. The forecasts verifying observations are subsequently matched and verified.

1.3.1 Verification Sites. A database of the NFDRS observation sites used in verification is posted to the NWS Performance Management website.

1.3.2 Verification Reports. NWS employees access verification statistics from the *Stats on Demand* feature of the NWS Performance Management website. Data are only available for the CONUS. *Stats on Demand* accesses an interactive database and generates verification statistics customized to the user's request. With each data request, the user provides the following definitions and boundaries:

a. Element. See section 1.3.3.

b. Beginning and ending dates. Specific months within a longer specified valid period may also be selected, e.g., select all June, July, and August data from the valid period January 1, 2004, thru December 31, 2008.

c. Spatial domain, to include (1) one or more individual verification sites, (2) one or more fire weather forecast zones, (3) one or more WFO forecast areas, or (4) the entire Nation, excluding Alaska.

d. Threshold error value. For temperature, relative humidity and wind speed, the user must include a threshold absolute error value. This value is entered by the user after selecting the desired forecast element (i.e., temperature, relative humidity, or wind speed) and is used to calculate the percentage of time the absolute error was greater than or equal to the user-specified value. Examples: 5ºF (temperature), 10% (relative humidity), 10 mph (wind speed).

e. Threshold window for Probability of Detection (POD), False Alarm Ratio (FAR), and Critical Success Index (CSI). For temperature, relative humidity, and wind speed, the user must specify the window of values, i.e., lowest and highest, from which the POD, FAR and CSI will be calculated. These values are entered by the user after selecting the desired forecast element (i.e., temperature, relative humidity, or wind speed). Examples: between 90 ºF and 120 ºF (temperature), between 30 and 100 mph (wind speed).

1.3.3 Elements.

a. State of Weather. Each state of weather is designated by a weather code (single integer) value from zero to 9. Each weather code is assigned to one of following three groups: group i (weather codes zero and 1), group ii (weather codes 2 and 3), and group iii (weather codes 4 thru 9). A forecast is counted as a hit if it falls within the same group as the observation.

b. Temperature.

c. Relative Humidity.

d. Wind Speed.

1.4 Red Flag Warnings (RFW). Perform RFW verification manually at each WFO with an RFW program.

1.4.1 Defining Events and Warnings. For verification purposes, an event is defined (a) when observations are queried in a given zone to determine if weather conditions meet or exceed the locally established warning criteria, and (b) when local land management personnel determine prior to warning issuance that the fuels meet or exceed the critical burning threshold. Each WFO and its local users determine the specific, unique weather criteria for issuance of a RFW in its area of responsibility. When observations are not available in a zone, the determination of an event should be based on the objective opinion of an experienced forecaster. Events are not determined by the number of fire starts or by querying users to determine if they feel you "hit" or "missed" warnings.

In summary, warnings are issued based on two factors: weather and fuel conditions. The former is determined by the forecaster, and the latter is determined by the user. The latter is determined in advance of the warning issuance and doesn't change when the verification is done.

1.4.2 Matching Warnings and Events and Performing Verification. Treat each fire weather zone as a separate verification area. Therefore, count a warning covering three zones as three warned areas or three warnings. Record warnings and events in separate databases. All listings in the event database must meet weather warning criteria. Warnings are verified based on whether the zone experienced locally-established weather warning criteria.

Count one verified warning and one warned event whenever an event meeting weather warning criteria occurs in a warned zone. Count one missed event if an event meeting weather warning criteria occurs in a zone with no warning. However, if weather warning criteria were met, but a warning was not issued because the users determined that the fuels were insufficient to warrant a warning, then a missed event is not recorded.

Count one unverified warning (or false alarm) for each warned zone that does not meet weather warning criteria.

The majority of RFWs include wind and humidity criteria or some index based on these parameters. However, in some areas, some warnings are issued due to the occurrence of dry lightning. These two types of events can exhibit big differences in lead time, Probability of Detection (POD), False Alarm Ratio (FAR), and Critical Success Index (CSI). Therefore, all offices doing RFW verification will verify RFWs three ways: first, by tracking and verifying all events; second, by tracking and verifying just wind/humidity events or their equivalent; and third, by tracking and verifying just dry lightning events. All three sets of verification must be sent to the regional fire weather program manager and/or verification program manager at the end of the calendar year. If an office does not have any criteria for dry lightning events, that office will report "n/a" for their dry lightning verification.

1.4.3 Extensions. Warnings may be extended in area and/or time. Count extensions of warnings to new areas (zones) as new warnings, i.e., one warning per zone.

1.4.4 Lead Time. Compute a lead time for each zone that experiences an event. Subtract the time of warning issuance from the time when the event first met warning criteria in the zone. If warning criteria at a particular WFO are subject to a temporal limit (e.g., the criteria must be met for a minimum of three consecutive hours), then the lead time is computed from the first observed occurrence of that temporal criteria. For example, a warning was issued at 0600 LST and the weather criteria were first met at 1200 LST. However, based upon the established temporal limit, the third hour of the weather criteria was not observed until 1500 LST. Assuming warning criteria as stated in the local Annual Operating Plan (AOP) have been met, then the calculated lead time would be 6 hours, i.e., 1200 LST (first occurrence) minus 0600 LST (warning issuance time).

Set negative values to zero. If a zone experiences an event meeting warning criteria when a warning is not in effect, assign that event a lead time of zero. Compute average lead time from all the lead times listed in the event database, including zeroes.

1.4.5 Regional Reports. The NWS regional headquarters report the annual verification statistics to the National Fire Weather Operations Coordinator (NFWOC). The report should contain the following elements by office: Number of RFWs issued, average lead time in hours, number of correct warnings, number of warnings that did not verify and number of unwarned events. These elements need to be reported for all events, just wind/humidity events or their equivalent, and just dry lighting events. Also include the number of spot forecasts issued by each office. The NFWOC will send the regional fire weather program managers a spreadsheet to fill in these numbers the first week of January. The NWS regions will report these numbers to the NFWOC by the last day in January. The NFWOC will compute the POD, FAR and CSI for each office, each region and nationally, as well as the average lead time for each region and nationally. POD, FAR and CSI are computed as follows:

a. Number of correct warnings (A)

b. Number of unwarned events (B)

c. Number of warnings that did not verify (C)

d. POD = A/(A+B)

e. FAR = C/(A+C)

f. CSI = A/(A+B+C)

1.5 Winter Weather Warnings. Automated verification of winter weather warnings is performed at the OCWWS Performance Branch.

a. NWS employees access these verification statistics through the *Stats on Demand* feature of the NWS Performance Management website. *Stats on Demand* uses an interactive database to provide verification statistics customized to the user's request. With each data request, the user provides the following definitions and boundaries:

(1) Type of warning (generic or one of the event-specific varieties listed in Table 1).

(2) Beginning and ending dates.

(3) One or more zones, WFOs, states, or NWS regions.

b. All winter weather warning verification is performed using one of the following methodologies. Advisories are not verified. The user of *Stats on Demand* must specify which method to use when requesting data. The default selection is "All Winter Events (Generic)."

(1) <u>All Winter Events (Generic)</u>. Any type of winter weather event that meets warning criteria verifies any type of winter weather warning, and any winter weather warning covers any winter event that meets warning criteria. See Table 1. This is the most frequently used method, and the method used in all Government Performance Results Act (GPRA) reports. It is also the default setting on the *Stats on Demand* winter weather warning request interface.

(2) <u>Event Specific</u>. Each warning must be verified with the exact, event specific *Storm Data* entry, e.g., an ice storm warning must be verified with an ice storm entry in *Storm Data* and vice versa. See Table 1.

Table 1. *Storm Data* entries (events), warning types, and verification methods.

<table>
<tr><th>Warning Type</th><th>Storm Data entries required for Event Specific verification of the warnings in column 1 and vice versa</th><th>Storm Data entries required for Generic verification of the warnings in column 1 and vice versa</th><th>Storm Data entries that do <u>not</u> verify the warnings in column 1 (warning criteria were not reached)</th></tr>
<tr><td>Winter Storm</td><td>Not applicable</td><td rowspan="4">Winter Storm,
Heavy Snow,
Sleet,
Ice Storm,
Lake Effect Snow,
or
Blizzard</td><td rowspan="4">Winter weather</td></tr>
<tr><td>Ice Storm</td><td>Ice Storm only</td></tr>
<tr><td>Lake Effect Snow</td><td>Lake Effect Snow only</td></tr>
<tr><td>Blizzard</td><td>Blizzard only</td></tr>
</table>

1.5.1 <u>Matching Warnings and Events</u>. All warning data are automatically taken from the warning products issued to the public. Each public forecast zone is treated as a separate verification area. Therefore, a warning covering three zones is counted as three warned areas or three warnings. All events that meet regional/local warning criteria (Table 1, columns 2 and 3) are automatically taken from the certified *Storm Data* reports.

The following event times, defined in NWSI 10-1605, Storm Data Preparation, are provided for each event listed in *Storm Data* and are used for verification:

a. Beginning time.

b. Criteria time.

c. Ending time.

Warnings and events that meet warning criteria are recorded in separate verification databases. Whenever the time period between the criteria time and the ending time of an event coincides with any part of the valid period of a warning, one warned event and one verified warning are counted. Unwarned events and unverified warnings are also counted. From these statistics, the POD, FAR, and CSI are computed (see appendix A, sections 3.1 to 3.3) and listed in the verification reports. Numerous examples of specific verification scenarios are provided in the "Winter Storm Warning Verification Overview" on the public forecast verification page of the NWS Performance Management website.

1.5.2 Quality Assurance. All data imported into the warning database are taken directly from the warning text. The issuing WFO and warning type in the VTEC line are checked for consistency with the warning header (top two lines of the warning). Inconsistent warnings are not counted for verification, and products issued with the improper coding may not be correctly imported into the database.

1.5.3 Extensions. Warnings may be extended in area and/or time. Extensions of warnings to new areas (zones) are counted as new warnings, i.e., one warning per zone. Each time extension of a zone already warned is counted as a new warning only if the earlier warning did not verify during its valid period. Examples of the verification of warning extensions are provided in the "Winter Storm Warning Verification Overview" on the public forecast verification page of the NWS Performance Management website.

1.5.4 Lead Time. A lead time is computed for each zone that experiences an event meeting warning criteria. If the event criteria time does not occur during the valid period of a warning, the lead time for that event is zero. If the event criteria time occurs during the valid period of a warning, the lead time for that event is computed by subtracting the warning issuance time from the event criteria time. Negative lead times are set to zero. The average lead time is computed from all lead times listed in the event database, including zeroes.

1.5.5 Timing Error. The timing error for each warned event is defined as the event beginning time minus the warning beginning time. For each data request, the mean absolute error and mean algebraic error (bias) are provided.

1.5.6 Watches and Advisories. While watches and advisories are not verified in the same manner as warnings, the following statistics are provided:

a. The percentage of unwarned events that occurred with an advisory in effect.

b. The percentage of unwarned events that occurred with a watch in effect.

1.5.7 Backup Mode for Warnings. All warnings issued by the backup office are attributed to the primary WFO, listed in the WMO (World Meteorological Organization) header of the warning.

1.6 High Wind Warnings. Automated verification of high wind warnings is performed at the OCWWS Performance Branch.

NWS employees access these verification statistics through the *Stats on Demand* feature of the NWS Performance Management website. *Stats on Demand* uses an interactive database to provide verification statistics customized to the user's request. With each data request, the user provides the following boundaries:

a. Beginning and ending dates.

b. One or more zones, WFOs, states, or NWS regions.

1.6.1 Matching Warnings and Events. All warning data are automatically taken from the warning products issued to the public. Each public forecast zone is treated as a separate verification area. Therefore, a warning covering three zones is counted as three warned areas or three warnings.

All events that meet warning criteria are automatically taken from certified *Storm Data* reports. The following event times, defined in NWSI 10-1605, Storm Data Preparation, are provided for each event listed in *Storm Data* and are used for verification:

a. Beginning time.

b. Ending time.

Warnings and events that meet warning criteria are recorded in separate verification databases. Whenever an event that meets warning criteria (defined temporally as the period between its beginning and ending times) coincides with any part of the valid period of a warning, one warned event and one verified warning are counted. Unwarned events and unverified warnings are also counted. From these tallied statistics, the POD, FAR, and CSI are computed (see appendix A, sections 3.1 to 3.3) and listed in the verification reports. Numerous examples of specific verification scenarios are provided in the "High Wind Warning Verification Overview" presentation on the public forecast verification page of the NWS Performance Management website.

1.6.2 Quality Assurance. All data imported into the warning database are taken directly from the warning text. The issuing WFO and warning type in the VTEC line are checked for consistency with the warning header (top two lines of the warning). Inconsistent warnings are not counted for verification, and products issued with the improper coding may not be correctly imported into the database.

1.6.3 Extensions. Warnings may be extended in area and/or time. Extensions of warnings to new areas (zones) are counted as new warnings, i.e., one warning per zone. Each time extension of a zone already warned is counted as a new warning only if the earlier warning did not verify during its valid period. Verification examples of warning extensions are provided in the "High Wind Warning Verification Overview" on the public forecast verification page of the NWS Performance Management website.

1.6.4 Lead Time. A lead time is computed for each zone that experiences an event meeting warning criteria. If the event beginning time does not occur during the valid period of a warning, the lead time for that event is zero. If the event beginning time occurs during the valid period of a warning, the lead time for that event is computed by subtracting the warning issuance time from the event beginning time. Negative lead times are set to zero. The average lead time is computed from all lead times listed in the event database, including zeroes.

1.6.5 Timing Error. The timing error for each warned event is defined as the event beginning time minus the warning beginning time. For each data request, the mean absolute error, the mean algebraic error (bias), and a distribution of errors are provided.

1.6.6 Watches and Advisories. While watches and advisories are not verified in the same manner as warnings, the following statistics are provided:

a. The percentage of unwarned events that occurred with an advisory in effect.

b. The percentage of unwarned events that occurred with a watch in effect.

1.6.7 Backup Mode for Warnings. All warnings issued by the backup office are attributed to the primary WFO, listed in the WMO header of the warning.

2. Convective Severe Weather Verification Procedures. This section describes the verification of all severe thunderstorm and tornado watches and warnings. The OCWWS Performance Branch is responsible for the operation and maintenance of the automated county-based and storm-based severe weather warning verification programs.

2.1 County-based Warning Verification. County-based warning issuance ceased October 1, 2007, so county-based warning verification should be used for warnings issued before this date. Storm-based warning issuance commenced on October 1, 2007; see section 2.2 for a description of storm-based warning verification. NWS employees access verification statistics through the *Stats on Demand* feature of the NWS Performance Management website. *Stats on Demand* uses an interactive database to provide verification statistics customized to the user's request. With each data request, the user provides the following definitions and boundaries:

a. Type of warning.

b. Beginning and ending dates.

c. One or more counties, WFOs, states, or NWS regions.

d. Severity of event, based on total cost of damage, number of fatalities, and/or tornado EF-scale (optional).

Verification statistics are computed for tornado and severe thunderstorm warnings and events using one of three methods. The user of *Stats on Demand* selects the method. The first method combines severe thunderstorms and tornadoes together and treats them as a single event type. The latter two methods are event specific—they treat non-tornadic severe thunderstorms and tornadoes as separate types of events. See Table 2 for illustration.

Table 2. *Storm Data* entries (events) used to verify local severe storm warnings.

Warning Type	**Event Specific Verification** Each warning type in the left column is only verified by the corresponding event type from the same line of this column. Each event type in this column must be covered by the corresponding warning type from the same line in the left column.	**All Severe Thunderstorm and Tornado (Generic) Verification** Each warning type in the left column is verified by *any* of the event types in this column. Any event type in this column must be covered by one of the warning types in the left column.
Severe thunderstorm (SVR product)	Non-tornadic severe thunderstorm, e.g., hail or thunderstorm wind meeting NWS warning criteria	Non-tornadic severe thunderstorm or tornado
Tornado (TOR product)	Tornado	

2.1.1 Matching Warnings and Events. All warning data are automatically extracted from the warning products issued to the public. Each county included in a warning statement is counted as a separate warning. The warning issuance and expiration times are taken from the Valid Time and Event Code (VTEC) line of the warning text.

All events are automatically taken from the certified *Storm Data* reports. Each warning (SVR or TOR) is only verified by a confirmed event of the type specified in Table 2. For verification purposes, multiple severe thunderstorm wind and hail events in the same county separated by less than 10 miles and 15 minutes are considered duplicates; only the first entry is used for verification. This rule has the following exceptions:

a. Any event that causes death or injury is included in the event database.

b. Any event that causes crop or property damage in excess of $500,000 is included in the event database.

c. Any report of winds 65 knots or greater is included in the event database.

d. Any hail size report of 2 inches or greater is included in the event database.

e. An event is not considered a duplicate if it is the only event verifying a warning.

Any event not recorded in the verification database due to the aforementioned duplicate rule may still appear in the publication *Storm Data*. An event moving into a second county creates an additional event for the database.

Warnings and events qualified for use in verification are recorded in separate databases. Whenever an event occurs in a warned county during any part of the valid period of the warning, one verified warning and one warned event are counted. Unwarned events and unverified warnings are also counted. From these databases, the POD, FAR, and CSI are computed (see appendix A, sections 3.1 to 3.3) and listed in the verification reports.

2.1.2 Quality Assurance. All data imported into the warning database are taken directly from the warning text. The issuing WFO and warning type in the VTEC line are checked for consistency with the warning header (top two lines of the warning). Inconsistent warnings are not counted for verification, and products issued with the improper coding may not be correctly imported into the database.

2.1.3 Lead Time. The methodologies for computing the lead time in each county for tornado, severe thunderstorm, and generic severe thunderstorm/tornado events are identical. For verification purposes, the definition of the term “event” is given in section 2.1.1. The lead time for each event is computed by subtracting the warning issuance time from the time when the event was first reported in the county. The warning issuance time is taken from VTEC line of the warning, and the start time of the event is taken from *Storm Data.* Negative lead times are set to zero. If one or more events occur in a county not covered by a warning, each unwarned event is assigned a lead time of zero. An event moving into a second county creates an additional event for the database. The lead time for the second event is based on the time the event first entered the second county. Average lead time is computed from all lead times listed in the event database, including zeroes. The percentage of events with a lead time greater than zero is also computed.

2.1.4 Backup Mode for Warnings. All warnings issued by the backup office are attributed to the primary WFO, listed in the WMO header of the warning.

2.2 Storm-based Warning Verification. Storm-based warning issuance replaced county-based warning issuance October 1, 2007, so storm-based warning verification should be used for warnings issued on or after this date. For warnings issued before October 1, 2007, see section 2.1 for a description of county-based warning verification. NWS employees access verification statistics through the *Stats on Demand* feature of the NWS Performance Management website. *Stats on Demand* uses an interactive database to provide verification statistics customized to the user's request. With each request, the user provides the following definitions and boundaries:

a. Type of warning (method).

b. Beginning and ending dates.

c. One or more WFOs, states, or NWS regions.

d. Severity of event, based on total cost of damage, number of fatalities, and/or tornado EF-scale (optional).

2.2.1 Matching Warnings and Events. All warning data are automatically extracted from the warning products issued to the public. The basic area for a tornado or severe thunderstorm warning is the polygon boundary outlined by the latitude-longitude coordinates located at the bottom of the product. Therefore, for verification purposes, the area within the latitude-longitude boundaries is counted as the warning.

Verification statistics are computed for tornado and severe thunderstorm warnings and events using one of three methods. The user of *Stats on Demand* selects the method. The first method combines severe thunderstorms and tornadoes together and treats them as a single event type. The latter two methods are event specific—they treat non-tornadic severe thunderstorms and tornadoes as separate types of events. See Table 2 for illustration.

All event data are automatically taken from the certified *Storm Data* reports. Each warning (SVR or TOR) is only verified by a confirmed event of the type specified in Table 2 and occurring within the temporal and areal boundaries of the warning. Unlike the county-based severe weather verification method, multiple severe thunderstorm wind and hail events in the same county, separated by less than 10 miles and 15 minutes, are ***not*** considered duplicates.

When categorizing a warning as verified or unverified, a check is executed to see if a verifying event occurred within the temporal and areal boundaries of that warning. If an event occurs within that warned area, one verified warning is recorded. If no event occurs within that warned area, an unverified warning is recorded.

Events are logged in *Storm Data* using one of two methods. The first method is an isolated event at a single location (referred to as an instantaneous event). An example would be an isolated hail event reported at a single time. The second method is used for an event that starts at one location and moves to another location over a period of time (referred to as a track event). An example would be a tornado that moved from one location to another. Both methods are evaluated differently.

a. Evaluation of Instantaneous Events. A check is performed on each instantaneous event to see if a warning was in effect at the time and location of the event. If so, the event was warned. If not, the event was unwarned.

b. Evaluation of Track Events. Before track events can be evaluated, two assumptions must be made:

(1) The event travels in a straight path between the event beginning and ending locations logged in *Storm Data.*

(2) The event travels at a constant speed between the event beginning and ending locations logged in *Storm Data.*

Once these assumptions are made, the location of the event is estimated every minute for the duration of the event. The event is then evaluated at each of those locations and times. For example, a tornado event lasting from 0100 to 0105 and traveling three miles would be evaluated at six locations and times. A check is then performed at each point along the track of the event to see if a warning was in effect. If so, the event was warned at that point. If not, the event was unwarned at that point.

From the event and warning databases, the POD, FAR, and CSI are computed (see appendix A, sections 3.1 to 3.3) and listed in the verification reports.

2.2.2 Quality Assurance. All data imported into the warning database are taken directly from the warning text. The issuing WFO and warning type in the VTEC line are checked for consistency with the warning header (top two lines of the warning). Inconsistent warnings are not counted for verification, and products issued with the improper coding may not be correctly imported into the database.

2.2.3 Lead Time. The methodologies for computing the lead time for tornado, severe thunderstorm, and generic severe thunderstorm/tornado events are identical. Lead time is calculated by subtracting the time of warning issuance from the time when an event was first reported in the warned area. The time of warning issuance is taken from the VTEC line of the warning, and the time when the event was first reported in the warned area is taken from *Storm Data*.

a. Lead Time of Instantaneous Events. A check is performed on instantaneous events to identify if a warning was valid at the time and location the event occurred. If a warning was valid at that time and location, the event is assigned a lead time based on the time the warning was issued. If no warning was valid at that time and location, the event is assigned a lead time of zero.

b. Lead Time of Track Events. The methodology of how track events are evaluated in one-minute points is given in section 2.2.1. A check is performed at each one-minute point, for the duration of the event, to identify if a warning was valid at the time and location along the path that the event occurred. If a warning was valid at the time and location, the point is assigned a lead time based on the time the warning was issued. If no warning was valid at that time and location, the point is assigned a lead time of zero. This process is repeated at every point for the duration of the event.

Event average lead times are assigned to each event in the detailed verification reports. These lead times are not used for the calculation of the overall average lead time for the entire verification report. Event average lead times are only created to show how much lead time was given for that event.

The overall average lead time, displayed in the summary of all reports, is calculated by averaging the lead times assigned to all points, both on instantaneous and track events. By calculating lead time via this method, long track events have more of an impact on the overall average lead time than instantaneous events.

The percentage of events with an event average lead time greater than zero is also computed.

2.2.4 Backup Mode for Warnings. All warnings issued by the backup office are attributed to the primary WFO, listed in the WMO header of the warning.

2.3 Watch Verification. The Storm Prediction Center (SPC) is responsible for verifying the tornado and severe thunderstorm watches it issues. The area enclosed by a watch is verified without regard to the number of counties affected. Weiss et al. (1980) describes how SPC accounts for variations in the size of convective watch areas. All event data are taken from the OCWWS database. Statistics are stratified for tornado and severe thunderstorm watches combined and for tornado watches only.

3. Marine Forecast Verification Procedures.

3.1 Coded Marine Forecasts.

3.1.1 Introduction. Marine wind and wave forecasts are verified at fixed point locations for specific time periods. The Ocean Prediction Center (OPC), Tropical Prediction Center (TPC), and WFOs with marine forecast responsibility will issue coded marine verification forecasts (MVF) twice a day for each verification site in their individual coastal waters (CWF), offshore

(OFF), Great Lake near shore (NSH), and Great Lake open lake (GLF) forecast areas. Discontinue the issuance of the MVF in the absence of operational verification sites within your area of responsibility. WFOs with marine responsibility are listed in Tables 3 and 4.

Table 3. Coastal WFOs with marine responsibility.

Eastern Region WFOs Caribou, ME (CAR) Portland, ME (GYX) Boston, MA (BOX) New York City, NY (OKX) Philadelphia, PA (PHI) Baltimore, MD/Washington DC (LWX) Wakefield, VA (AKQ) Morehead City, NC (MHX) Wilmington, NC (ILM) Charleston, SC (CHS) Southern Region WFOs Jacksonville, FL (JAX) Melbourne, FL (MLB) Miami, FL (MFL) Key West, FL (EYW) Tampa Bay Area, FL (TBW) Tallahassee, FL (TAE) Mobile, AL (MOB) New Orleans, LA (LIX) Lake Charles, LA (LCH) Houston/Galveston, TX (HGX) Corpus Christi, TX (CRP) Brownsville, TX (BRO) San Juan, PR (TJSJ)	Western Region WFOs Seattle, WA (SEW) Portland, OR (PQR) Medford, OR (MFR) Eureka, CA (EKA) San Francisco, CA (MTR) Los Angeles, CA (LOX) San Diego, CA (SGX) Alaska Region WFOs Juneau, AK (PAJK) Anchorage, AK (PAFC) Fairbanks, AK (PAFG) Pacific Region WFOs Honolulu, HI (PHFO) Guam (PGUM) Pago Pago (NSTU)

Table 4. Great Lakes WFOs with marine responsibility

Eastern Region WFOs Cleveland, OH (CLE) Buffalo, NY (BUF)	Central Region WFOs Duluth, MN (DLH) Marquette, MI (MQT) Gaylord, MI (APX) Detroit, MI (DTX) Green Bay, WI (GRB) Milwaukee, WI (MKX) Chicago, IL (LOT) Grand Rapids, MI (GRR) Northern Indiana (IWX)

3.1.2 Verification Sites. The WFOs with marine responsibility, OPC, and TPC will use any reliably functioning buoy or Coastal Marine Automated Network (C-MAN) station residing within their respective forecast areas as a verification site. Remove any buoy or C-MAN that becomes unreliable or inactive, i.e., no data available for verification. WFOs with Great Lakes marine responsibility will discontinue the MVF after the buoys are removed from the lakes for the winter. An interactive directory of national marine verification sites appears on the marine forecast verification page of the NWS Performance Management website.

3.1.3 Coded Forecast Format. Code the MVF in accordance with the format in Table 5. Issue the MVF no later than 2 hours after issuing the CWF, OFF, NSH, or GLF, using forecast values meteorologically consistent with the worded forecasts, remembering the winds and waves in the

Table 5. Definitions of code used in the MVF. See text for detailed explanation.

CODE FORMAT *%%F nn(space)xxxxx(space)t_1t_1/WW/ddff/hh/t_2t_2/WW/ddff/hh [LF][LF]$$*	
%%F	Code for computer and delimiter for operational forecast
nn	Forecaster number
xxxxx	Buoy/C-MAN identifier
t_1t_1	Time, in hours (UTC), of the midpoint of the valid period for the 16- to 20-hour forecast, i.e., 06 or 18 UTC.
WW	Warning/advisory status NO: No advisory or warning SC: Small craft advisory GL: Gale warning ST: Storm warning TS: Tropical storm warning HR: Hurricane warning HF: Warning for hurricane force winds in the absence of a hurricane
dd	Wind direction
ff	Wind speed
hh	Significant wave height
t_2t_2	Time, in hours (UTC), of the midpoint of the valid period for the 28- to 32-hour forecast, i.e., 06 or 18 UTC.
[LF][LF]$$	End bulletin code (2 line feeds followed by turn off code)

Table 6. Examples of marine products.

Example of a segment of a Coastal Waters Forecast:

PZZ150-153-156-170-173-176-221715-
/X.EXT.KSEW.GL.W.0002.000000T0000Z-050523T0100Z/
/X.EXB.KSEW.SC.Y.0013.050523T0100Z-050524T0100Z/
CAPE FLATTERY TO CAPE SHOALWATER OUT TO 60 NM-
242 AM PDT SUN MAY 22 2005

...GALE WARNING EXTENDED UNTIL THIS AFTERNOON...
...SMALL CRAFT ADVISORY IN EFFECT FROM THIS EVENING TO MONDAY AFTERNOON...

.TODAY...SW WIND 25 TO 35 KT. WIND WAVES 4 TO 6 FT. SW SWELL 15 FT AT 9 SECONDS...BUILDING TO 21 FT AT 12 SECONDS. SHOWERS LIKELY.
.TONIGHT...SW WIND 20 TO 25 KT...EASING TO 15 KT AFTER MIDNIGHT. WIND WAVES 2 TO 4 FT. W SWELL 17 FT AT 11 SECONDS. CHANCE OF SHOWERS.

(rest of CWF text)

Example of Corresponding Coded MVF:

FXUS56 KSEW 221030
MVF001

%%F56 46041 18/GL/2235/21/06/SC/2623/17
$$

MVF are intended only for the sensors of the buoys and C-MAN stations. See Table 6 for a sample CWF with the corresponding MVF. A detailed explanation for each MVF entry is given below:

a. Forecaster Number (*nn*). Each forecaster is assigned a number.

b. Buoy/C-MAN Identifier (*xxxxx*). See section 3.1.2, Verification Sites.

c. Valid Periods (t_1t_1 and t_2t_2). The first valid period (t_1t_1) (UTC) in the MVF is 18 hours ± 2 hours following the 0000 or 1200 UTC model cycle, i.e., 1600 to 2000 UTC today for today's 0000 UTC cycle and 0400 to 0800 UTC tomorrow for today's 1200 UTC cycle. Therefore, the *WW*, *dd*, *ff*, and *hh* values immediately after t_1t_1 are 16- to 20-hour forecasts. The second valid period (t_2t_2) in the MVF is 30 hours ± 2 hours following the 0000 or 1200 UTC model cycle, i.e., 0400 to 0800 UTC tomorrow for today's 0000 UTC cycle and 1600 to 2000 UTC tomorrow for today's 1200 UTC cycle. Therefore, the *WW*, *dd*, *ff*, and *hh* values immediately after t_2t_2 are 28- to 32-hour forecasts.

d. Warning/Advisory Status for Wind and Waves (*WW*). Each entry results from the highest of five hourly sustained 10-meter wind speed forecasts or the highest of five hourly significant wave height forecasts for each of the MVF valid periods. Each hourly sustained wind speed is defined as the 10-minute mean wind speed corrected to 10 meters above sea level. The 2-character entry and threshold values used to define each warning/advisory category follow:

(1) NO: No warning or small craft advisory. When sustained wind speed is not forecast, "NO" is entered as the placeholder. When a small craft advisory is issued in the near-shore forecast solely for waves (Great Lakes only), "NO" is entered into the MVF because C-MAN stations do not measure wave height.

(2) SC: Small craft advisories. These are issued only for coastal water forecasts and near shore (Great Lake) forecasts. They are not issued for offshore or open lake forecasts. The sustained wind speed and significant wave height minimum threshold values used to define Small Craft Advisories are set locally or regionally and are programmed into the marine verification software. A forecast equaling or exceeding either the wind or wave minimum threshold value for a small craft advisory triggers the advisory.

(3) GL: Gale warning, sustained wind speeds 34 to 47 knots.

(4) ST: Storm warning, sustained wind speeds 48 to 63 knots.

(5) TS: Tropical storm warning, sustained wind speeds 34 to 63 knots in the presence of a tropical storm.

(6) HR: Hurricane warning, sustained wind speeds exceeding 63 knots in the presence of a hurricane.

(7) HF: Warning for hurricane force winds (sustained speeds exceeding 63 knots) in the absence of a hurricane.

If a gale warning is headlined in the offshore forecast due to forecast winds increasing to gale force during the 28- to 32-hour MVF valid period, "GL" is not entered as the advisory/warning category for both MVF valid periods. "NO" advisory/warning is entered for the 16- to 20-hour valid period, and "GL" advisory/warning is entered for the 28- to 32-hour valid period.

An MVF may contain a certain advisory/warning category in the MVF (e.g., gales), but simultaneously forecast a sustained speed (section 3.1.3.f, sustained wind speed, *ff*) less than the minimum threshold for that warning category. Both entries are probably legitimate because gale warnings are issued for the <u>maximum</u> forecast sustained speeds during the valid period, and wind speed is a forecast of the mean sustained wind speed expected during the valid period.

e. Wind Direction (*dd*). Wind direction forecasts are the unit vector resultant of the five hourly sustained wind direction forecasts for each MVF valid period. Each hourly sustained wind direction is defined as the forecast 10-minute mean wind direction corrected to 10 meters above sea level. The forecast is entered in tens of degrees, e.g., "12" represents a wind from 120 degrees. If a wind shift or variable winds are expected during the period, forecast direction at the midpoint hour of the valid period (i.e., 0600 or 1800 UTC) is entered. If the wind direction is less than 100 degrees, a zero is placed in the tens digit, e.g., "07" is entered for a wind from 70 degrees. When the wind speed equals 100 knots or more, 50 is added to wind direction, e.g., "57" is entered for a 100-knot or greater wind from 70 degrees. A "99" is entered if wind is forecast to be variable based on regional guidelines or the wind direction is not forecast due to missing observation data. See Table 7 for more examples.

Table 7. Examples of wind direction coded entries to the MVF.

For wind speeds less than 100 knots:		For speeds equal to or greater than 100 knots:	
Direction (degrees)	**Code**	**Direction (degrees)**	**Code**
Variable	99		
010	01	010	51
020	02	020	52
030	03	030	53
(and so on...)		(and so on...)	
300	30	300	80
310	31	310	81
320	32	320	82
(and so on...)		(and so on...)	

f. Wind Speed (*ff*). Wind speed forecasts are taken from the mean of the five hourly sustained wind speed forecasts for each MVF valid period. Enter the mean 10-meter sustained (10-minute mean) wind speed forecast for the valid period to the nearest knot. The value “99” is not entered. If the forecast wind speed is less than 10 knots, a zero is entered in the tens digit place, e.g., "06" is entered for 6 knots. For 100-knot or greater forecast speeds, 100 is subtracted from the forecast speed and 50 is added to the forecast direction. For example, given a forecast 110 knot wind from 270 degrees, the value “77” is entered for wind direction, and “10” is entered for wind speed. See paragraph e. and Table 7 for more details on wind direction. If the wind speed is not forecast due to missing observation data, enter "99" for wind speed; also enter "NO" as the placeholder in the warning/advisory position.

g. Significant Wave Height (*hh*). Significant wave height forecasts for each valid period are entered in whole feet. If less than 10 feet, a zero is placed in the tens digit, e.g., "08" is entered for 8 feet. If the significant wave height is not forecast due to missing or non-existent observation data (e.g., CMAN sites that do not measure significant wave height) "99" is entered as the placeholder.

3.1.4 Computation and Display of Verification Statistics. The OCWWS Performance Branch is responsible for operation and maintenance of the automated wind and wave marine verification program. NWS employees access verification statistics through the *Stats on Demand* feature of the NWS Performance Management website. *Stats on Demand* uses an interactive database to provide verification statistics customized to the user's request. With each request, the user provides the following boundaries:

a. Months.

b. Model cycles (0000 UTC for the early morning forecast; 1200 UTC for late afternoon).

c. Projections (18 or 30 hours).

d. Verification sites (single site, multiple sites, WFO area, regional data, national data).

e. Matching guidance product.

3.1.5 Verification Statistics. Verification statistics are computed for warning/advisory category, wind direction, wind speed and significant wave height. These statistics are based on a series of five hourly buoy or C-MAN observations within the MVF valid periods. A summary of each element follows.

a. Warning/Advisory Status for Wind and Waves. The warning/advisory status is verified against the highest of the five hourly wind speed observations during the MVF valid period.

(1) The lower sustained wind speed and significant wave height threshold values that define small craft advisories (SCA) are set locally or regionally, and these values are programmed into the marine verification software. Either the observed lower significant wave height threshold for an SCA or the observed lower sustained wind speed threshold for an SCA verifies the advisory. A 33-knot observed wind is the upper sustained wind speed threshold for verifying an SCA. SCAs are only issued for CWFs and NSHs.

(2) A 34- to 47-knot sustained wind speed verifies a gale warning.

(3) A 48- to 63-knot sustained wind speed verifies a storm warning.

(4) A 34- to 63-knot sustained wind speed verifies a tropical storm warning.

(5) A wind speed exceeding 63 knots verifies a hurricane warning or a warning for hurricane force winds in the absence of a hurricane.

The advisory/warning categories in the CWFs and NSHs are verified in 5x5 contingency tables of forecast categories versus observed categories. The warning categories in the OFFs and GLFs are verified in 4x4 contingency tables of forecast categories versus observed categories. From the contingency tables, the following statistics are computed: percent hits and the percentage of times that the forecast category missed the observed category by more than one category in either direction (i.e, positive and negative errors). POD, FAR, and CSI are also computed for gale warnings and greater.

b. Wind Speed. The coded forecast to the nearest knot is verified against the mean of the five hourly sustained wind speed observations during the MVF valid period. The observation sites used in verification may vary considerably in height and are corrected to the 10 meter standard forecast height by the National Data Buoy Center (NDBC) using Liu et al. 1979.

(1) Verification statistics are computed from the information contained in 7x7 contingency tables of forecasts versus observations. The wind speed categories are:

- Less than 8 knots.
- 8 to 12 knots.
- 13 to 17 knots.
- 18 to 22 knots.
- 23 to 27 knots.
- 28 to 32 knots.
- Greater than 32 knots.

(2) From the 7x7 contingency table, the following scores are computed: Gerrity Equitable Skill Score (ESS), percent hits (the percentage of time

that the forecast category equaled the observed category) (also called percent correct), and the percentage of time that the forecast category missed the observed category by more than one category. A high-end delta for the ESS is also computed. The POD, FAR, and CSI are also computed for the following wind speed thresholds: greater than 27 knots and greater than 32 knots.

The data are sorted by observation category (using the aforementioned categories), and the following statistics are provided: mean absolute error, mean algebraic error, and root mean square error. These statistics are also provided for the entire dataset regardless of observation categories.

(3) A histogram of absolute error frequencies (listed as percentages) is provided for each of the following error categories:

- 0 to 2 knots (over- and under-forecasting combined).
- 3 to 4 knots (over- and under-forecasting separate).
- 5 to 7 knots (over- and under-forecasting separate).
- 8 to 12 knots (over- and under-forecasting separate).
- 13 to 17 knots (over- and under-forecasting separate).
- 18 to 22 knots (over- and under-forecasting separate).
- Greater than 22 knots (over- and under-forecasting separate).

A separate error-based percent correct score, independent of the 7x7 contingency table, is also computed. It is defined as the percentage of forecasts with an error less than 5 knots (the sum of the first two bullets).

c. <u>Wind Direction</u>. Variable forecasts (coded '99') are not verified. Each forecast is verified with a time-averaged observation from the valid period of the MVF, omitting any observation with a reported wind speed less than 8 knots. The observations used in verification are corrected to the 10 meter standard forecast height by NDBC (Liu et al. 1979). Under most circumstances, this is the unit vector resultant of the five hourly reported directions during the forecast valid period. If any of the remaining 8-knot or greater winds varied in direction from any of the others in the valid period by more than 90 degrees, then the forecast is verified with the wind direction at the midpoint hour of the valid period, i.e., 0600 or 1800 UTC. If that midpoint hour wind speed was less than 8 knots and the reported directions varied by more than 90 degrees, then wind direction for that valid period is not verified.

(1) Verification statistics are computed from the information contained in 8x8 contingency tables of forecasts vs. observations. Each category is defined as one of the eight points of the compass:

- North: 338 to 22 degrees.
- Northeast: 23 to 67 degrees.
- East: 68 to 112 degrees.
- Southeast: 113 to 157 degrees.
- South: 158 to 202 degrees.
- Southwest: 203 to 247 degrees.
- West: 248 to 292 degrees.
- Northwest: 293 to 337 degrees.

(2) From the 8x8 contingency table, the following scores are computed: percent hits (the percentage of time that the forecast category equaled the observed category), and percentage of time that the forecast category missed the observed category by more than one category.

The data are sorted by observation category (using the aforementioned categories), and the following statistics are provided: mean absolute error, mean algebraic error, and root mean square error. These statistics are also provided for the entire dataset regardless of observation category.

(3) A histogram of absolute error frequencies (listed as percentages of the total) is provided for each of the following error categories:

- Hits.
- 1 octal error (over- and under-forecasting separate).
- 2 octal errors (over- and under-forecasting separate.
- 3 octal errors (over- and under-forecasting separate).
- 4 octal errors (over- and under-forecasting separate).

d. Significant Wave Height. The coded forecast to the nearest foot is verified against the mean of the five hourly significant wave height observations during the MVF valid period.

(1) Verification statistics are computed from the information contained in 7x7 contingency tables of forecasts versus observations. The categories are:

- Less than 3 feet.
- 3 to 5 feet.
- 6 to 8 feet.
- 9 to 12 feet.
- 13 to 16 feet.
- 17 to 20 feet.
- Greater than 20 feet.

(2) From the 7x7 contingency table, the following scores are computed: Gerrity Equitable Skill Score (ESS), percent hits (the percentage of time that the forecast category equaled the observed category) (also called percent correct), and percentage of time that the forecast category missed the observed category by more than one category. A high-end delta for the ESS is also computed. POD, FAR, and CSI are also computed for the following wind speed thresholds: greater than 27 knots and greater than 32 knots.

The data are sorted by observation category (using the aforementioned categories), and the following statistics are provided: mean absolute error, mean algebraic error, and root mean square error. These statistics are also provided for the entire dataset regardless of observation category.

(3) A histogram of absolute error frequencies (listed as percentages) are provided for each of the following error categories:

- Less than 2 feet (over- and under-forecasting combined).
- 2 to 3 feet (over- and under-forecasting separate).
- 4 to 5 feet (over- and under-forecasting separate).

- 6 to 8 feet (over- and under-forecasting separate).

- Greater than 8 feet (over- and under-forecasting separate).

A separate error-based percent correct score, independent of the 7x7 contingency table, is also computed. It is defined as the percentage of forecasts with an error less than 2 feet (i.e., also the first bullet in the histogram of error frequencies).

3.2 Coastal Flood and Lakeshore Flood Warnings (CFW). Automated verification of CFWs is performed at the OCWWS Performance Branch.

NWS employees access these verification statistics through the *Stats on Demand* feature of the NWS Performance Management website. *Stats on Demand* uses an interactive database to provide verification statistics customized to the user's request. With each request, the user provides the following boundaries:

a. Beginning and ending dates.

b. One or more zones, WFOs, states, or NWS regions.

3.2.1 Matching Warnings and Events. All warning data are automatically taken from the warning products issued to the public. Each public forecast zone is treated as a separate verification area. Therefore, a warning covering three zones is counted as three warned areas or three warnings.

a. All events are automatically taken from the certified *Storm Data* reports. Only the following reportable events are used to verify a CFW:

   (1) Coastal Flood.

   (2) Lakeshore Flood.

   (3) Seiche.

b. See NWSI 10-1605, Storm Data Preparation, for descriptions of each of the above events. Minor coastal or lakeshore flooding, such as nuisance flooding, is treated as a non-event for verification purposes. The following event times, defined in NWSI 10-1605, Storm Data Preparation, are provided for each event listed in *Storm Data* and are used in verification:

   (1) Beginning time.

   (2) Ending time.

c. Warnings and reportable events are recorded in separate verification databases. Whenever a reportable event (defined as the period between its beginning and ending times) coincides with any part of the valid period of a warning, one warned event and one verified warning are counted. Unwarned events and unverified warnings are also counted. From these databases, the POD, FAR, and CSI are computed (see appendix A, sections 3.1 to 3.3) and listed in the verification reports. Numerous examples of verification scenarios are provided in the "Coastal Flood Warning Verification Overview" presentation on the marine forecast verification page of the NWS Performance Management website.

3.2.2 Quality Assurance. All data imported into the warning database are taken directly from the warning text. The issuing WFO and warning type in the VTEC line are checked for consistency with the warning header (top two lines of the warning). Inconsistent warnings are not counted for verification, and products issued with the improper coding may not be correctly imported into the database.

3.2.3 Extensions. Warnings may be extended in area and/or time. Extensions of warnings to new areas (zones) are counted as new warnings, i.e., one warning per zone. Each time extension of a zone already warned is counted as a new warning only if the earlier warning did not verify during its valid period. Verification examples of warning extensions are provided in the "Coastal Flood Warning Verification Overview" on the marine forecast verification page of the NWS Performance Management website.

3.2.4 Lead Time. A lead time is computed for each zone that experiences a reportable event. If the event beginning time does not occur during the valid period of a warning, the lead time for that event is zero. If the event beginning time occurs during the valid period of a warning, the lead time for that event is computed by subtracting the warning issuance time from the event beginning time. Negative lead times are set to zero. The average lead time is computed from all lead times listed in the event database, including zeroes.

3.2.5 Timing Error. The timing error for each warned event is defined as the event beginning time minus the warning beginning time. For each data request, the mean absolute error, the mean algebraic error (bias), and a distribution of errors are provided.

3.2.6 Watches. While watches are not verified in the same manner as warnings, the percentage of unwarned events that occurred with a watch in effect is provided.

3.2.7 Backup Mode for Warnings. All warnings issued by the backup office are attributed to the primary WFO, listed in the WMO header of the warning.

3.3 Zone-Based Special Marine Warning (SMW) Verification. The OCWWS Performance Branch operates and maintains the automated marine zone-based SMW verification program. Any SMW issued for a coastal or Great Lake marine zone, Lake Okeechobee, or Lake

Pontchartrain is verified. Marine zone-based SMW issuance ceased October 1, 2007, so marine zone-based SMW verification should be used for warnings issued before this date. Storm-based SMW issuance commenced on October 1, 2007; see section 3.4 for a description of storm-based SMW verification.

NWS employees access these verification statistics through the *Stats on Demand* feature of the NWS Performance Management website. *Stats on Demand* uses an interactive database to provide verification statistics customized to the user's request. With each data request, the user provides the following definitions and boundaries:

a. Beginning and ending dates.

b. One or more marine zones, WFOs, bodies of water, or NWS regions.

c. Severity of event, based on total cost of damage and/or number of fatalities (optional).

3.3.1 Matching Warnings and Events. All warning data are automatically taken from the warning products issued to the public. Each marine forecast zone represents a separate verification area. Therefore, a warning issued for two zones counts as two separate warnings. Only the following reportable events in the certified *Storm Data* reports verify the SMW:

a. Marine hail, 3/4 inch or greater.

b. Marine thunderstorm wind, 34 knots or greater.

c. Waterspouts.

d. Marine strong wind.

e. Marine high wind.

Warnings and reportable events are recorded in separate databases. Whenever a reportable event occurs in a warned marine zone during any part of the valid period of the warning, one verified warning and one warned event are counted. Unwarned events and unverified warnings are also counted. From these databases, the POD, FAR, and CSI are computed (see appendix A, sections 3.1 to 3.3) and listed in the verification reports.

3.3.2 Quality Assurance. All data imported into the warning database are taken directly from the warning text. The issuing WFO and warning type in the VTEC line are checked for consistency with the warning header (top two lines of the warning). Inconsistent warnings are not counted for verification, and products issued with the improper coding may not be correctly imported into the database.

3.3.3 Lead Time. The lead time for each reportable event is computed separately for each marine zone by subtracting the time of warning issuance from the time when the reportable event was first reported in the marine zone. The time of warning issuance is taken from the VTEC line, and the time when the reportable event was first reported in the marine zone is taken from *Storm Data*. Negative lead times are set to zero. If one or more events occur in a zone with no warning in effect, each unwarned event is assigned a lead time of zero. Average lead time is computed from all lead times listed in the event database, including the zeroes. The percentage of events with a lead time greater than zero is also computed and listed in the verification reports.

3.3.4 Backup Mode for Warnings. All warnings issued by the backup office are attributed to the primary WFO, listed in the WMO header of the warning.

3.4 Storm-Based SMW Verification. The OCWWS Performance Branch operates and maintains the automated storm-based SMW verification program. Any SMW issued for a coastal or Great Lake marine zone, Lake Okeechobee, or Lake Pontchartrain is verified. Storm-based SMW issuance replaced marine zone-based SMW issuance October 1, 2007, so storm-based SMW verification should be used for warnings issued on or after this date. For warnings issued before October 1, 2007, see section 3.3 for a description of marine zone-based SMW verification.

NWS employees access these verification statistics through the *Stats on Demand* feature of the NWS Performance Management website. *Stats on Demand* uses an interactive database to provide verification statistics customized to the user's request. With each request, the user provides the following definitions and boundaries:

a. Beginning and ending dates.

b. One or more WFOs or NWS regions.

c. Severity of event, based on total cost of damage and/or number of fatalities (optional).

3.4.1 Matching Warnings and Events. All warning data are automatically extracted from the warning products issued to the public. The basic area for a SMW is the polygon boundary outlined by the latitude-longitude coordinates located at the bottom of the product. Therefore, for verification purposes, the area within the latitude-longitude boundaries is counted as the warning.

a. Only the following reportable events in the certified *Storm Data* reports occurring within the temporal and areal boundaries of an SMW verify that warning:

(1) Marine hail, 3/4 inch or greater.

(2) Marine thunderstorm wind, 34 knots or greater.

(3) Waterspouts.

(4) Marine strong wind.

(5) Marine high wind.

b. When categorizing a warning as verified or unverified, a check is executed to see if a verifying event occurred within the temporal and areal boundaries of that warning. If an event occurs within that warned area, one verified warning is recorded. If no event occurs within that warned area, an unverified warning is recorded.

Events are logged in *Storm Data* using one of two methods. The first method is an isolated event at a single location (referred to as an instantaneous event). An example would be an isolated marine hail event reported at a single time. The second method is used for an event that starts at one location and moves to another location over a period of time (referred to as a track event). An example would be a waterspout that moved from one location to another. Both methods are evaluated differently.

(1) <u>Evaluation of Instantaneous Events.</u> A check is performed on each instantaneous event to see if a warning was in effect at the time and location of the event. If so, the event was warned. If not, the event was unwarned.

(2) <u>Evaluation of Track Events.</u> Before track events can be evaluated, two assumptions must be made:

- The event travels in a straight path between the event beginning and ending locations logged in *Storm Data.*

- The event travels at a constant speed between the event beginning and ending locations logged in *Storm Data.*

Once these assumptions are made, the location of the event is estimated every minute for the duration of the event. The event is then evaluated at each of those locations and times. For example, a waterspout event lasting from 0100 to 0105 and traveling three miles would be evaluated at six locations and times. A check is then performed at each point along the track of the event to see if a warning was in effect. If so, the event was warned at that point. If not, the event was unwarned at that point.

c. From the event and warning databases, the POD, FAR, and CSI are computed (see appendix A, sections 3.1 to 3.3) and listed in the verification reports.

3.4.2 Quality Assurance. All data imported into the warning database are taken directly from the warning text. The issuing WFO and warning type in the VTEC line are checked for consistency with the warning header (top two lines of the warning). Inconsistent warnings are not counted for verification, and products issued with the improper coding may not be correctly imported into the database.

3.4.3 Lead Time. Lead time is calculated by subtracting the time of warning issuance from the time when a reportable event first occurred in the warned area. The time of warning issuance is taken from the VTEC line of the warning, and the time when a reportable event first occurred in the warned area is taken from *Storm Data*.

a. Lead Time of Instantaneous Events. A check is performed on instantaneous events to identify if a warning was valid at the time and location the event occurred. If a warning was valid at that time and location, the event is assigned a lead time based on the time the warning was issued. If no warning was valid at that time and location, the event is assigned a lead time of zero.

b. Lead Time of Track Events. The methodology of how track events are evaluated in one-minute points is given in section 3.4.1. A check is performed at each one-minute point, for the duration of the event, to identify if a warning was valid at the time and location along the path that the event occurred. If a warning was valid at the time and location, the point is assigned a lead time based on the time the warning was issued. If no warning was valid at that time and location, the point is assigned a lead time of zero. This process is repeated at every point for the duration of the event.

Event average lead times are assigned to each event in the detailed verification reports. These lead times are not used for the calculation of the overall average lead time for the entire verification report. Event average lead times are only created to show how much lead time was given for that event.

The overall average lead time score, displayed in the summary of all reports, is calculated by averaging the lead times assigned to all points, both on instantaneous and track events. By calculating lead time via this method, long track events have more of an impact on the average lead time score than instantaneous events.

The percentage of events with an event average lead time greater than zero is also computed.

3.4.4 Backup Mode for Warnings. All warnings issued by the backup office are attributed to the primary WFO, listed in the WMO header of the warning.

4. Hydrologic Verification Procedures. Hydrologic verification consists of the verification of county-based and storm-based flash flood warnings (FFW) and River Forecast Center (RFC) river stage forecasts.

4.1 County-Based FFW Verification. The OCWWS Performance Branch is responsible for the operation and maintenance of the automated county-based FFW verification program. See section 4.2 for a description of storm-based FFW verification.

NWS employees access these verification statistics through the *Stats on Demand* feature of the NWS Performance Management website. *Stats on Demand* uses an interactive database to provide verification statistics customized to the user's request. With each request, the user provides the following definitions and boundaries:

a. Beginning and ending dates.

b. One or more counties, WFOs, states, NWS regions, or the contiguous United States.

c. Severity of event, based on total cost of damage and/or number of fatalities (optional).

4.1.1 Matching Warnings and Events. All warning data are automatically extracted from the warning products issued to the public. Since each county specified in a warning represents a separate verification area, a warning covering three counties is counted as three warnings. Events are automatically taken from the certified *Storm Data* reports. *Storm Data* reports entered as the event type "flash flood" verify an FFW.

Warnings and events are recorded in separate databases. Whenever an event occurs in a warned county during any part of the valid period of the warning, one verified warning and one warned event are counted. Unwarned events and unverified warnings are also recorded and tallied. From these databases, the POD, FAR, and CSI are computed (see appendix A, sections 3.1 to 3.3) and listed in the verification reports.

4.1.2 Quality Assurance. All data imported into the warning database are taken directly from the warning text. The issuing WFO and warning type in the VTEC line are checked for consistency with the warning header (top two lines of the warning). Inconsistent warnings are not counted for verification, and products issued with the improper coding may not be correctly imported into the database.

4.1.3 Lead Time. For verification purposes, the definition of the term "event" is given in section 4.1.1. The lead time for each flash flood event is computed separately for each county by subtracting the time of warning issuance from the time when the event first occurred in the county. The time of warning issuance comes from the VTEC line in the FFW, and the event beginning time for the given county is taken from *Storm Data*. Negative lead times are set to zero. If one or more events occur in a county with no warning in effect, each unwarned event is assigned a lead time of zero. Average lead time is computed from all lead times listed in the event database, including zeroes. The percentage of events with lead time greater than zero is also computed.

4.1.4 Backup Mode for Warnings. All warnings issued by the backup office are attributed to the primary WFO, listed in WMO header of the warning.

4.2 Storm-Based FFW Verification. The OCWWS Performance Branch is responsible for the operation and maintenance of the automated storm-based FFW verification program. See section 4.1 for a description of county-based FFW verification.

NWS employees access these verification statistics through the *Stats on Demand* feature of the NWS Performance Management website. *Stats on Demand* uses an interactive database to provide verification statistics customized to the user's request. With each request, the user provides the following definitions and boundaries:

a. Beginning and ending dates.

b. One or more WFOs or NWS regions.

c. Severity of event, based on total cost of damage and/or number of fatalities (optional).

4.2.1 Matching Warnings and Events. All warning data are automatically extracted from the warning products issued to the public. The basic area for a FFW is the polygon boundary outlined by the latitude-longitude coordinates located at the bottom of the product. Therefore, for verification purposes, the area within the latitude-longitude boundaries is counted as the warning.

All event data are automatically taken from certified *Storm Data* reports of the event type "flash flood." Each FFW may only be verified by a confirmed event occurring within the temporal and areal boundaries of the warning.

When categorizing a warning as verified or unverified, a check is executed to see if a verifying event occurred within the temporal and areal boundaries of that warning. If an event occurs within that warned area, one verified warning is recorded. If no event occurs within that warned area, an unverified warning is recorded.

From the event and warning databases, the POD, FAR, and CSI are computed (see appendix A, sections 3.1 to 3.3) and listed in the verification reports.

Events are logged in *Storm Data* as areal events, which means that each event is entered for an area of land. The area of the event reported and the forecast area of the warning are overlaid to compute the ***percentage of the event warned.***

4.2.2 Quality Assurance. All data imported into the warning database are taken directly from the warning text. The issuing WFO and warning type in the VTEC line are checked for consistency with the warning header (top two lines of the warning). Inconsistent warnings are not counted for verification, and products issued with the improper coding may not be correctly imported into the database.

4.2.3 Lead Time. The lead time for flash floods will be referred to as the ***maximum event lead time.*** The verification matching process will generate this maximum event lead time for all flash flood events. This score will be calculated by analyzing every flash flood event to see if a warning was valid over any area of the event when the event first began. Any amount of areal overlap of the warning and event polygons is acceptable. The key is to use only those warnings valid for the time and area when the event first begins.

There are three possible scenarios:

a. No warning is valid for the area in which an event begins. The Maximum Event Lead Time equals zero 0 minutes. This also applies to situations where warnings are issued after the event has already begun.

b. A single warning is valid for the area in which an event begins. The Maximum Event Lead Time equals the Event Beginning time minus the Warning Issuance Time.

c. Multiple warnings are valid for the area in which an event begins. In this case, the lead time calculation is based only on the warning that was issued earliest (first). The Maximum Event Lead Time equals the Event Beginning Time minus the Warning Issuance Time (of first issued warning).

The time of warning issuance is taken from the VTEC line of the warning, and the event beginning time is taken from *Storm Data.* Negative lead times are set to zero.

If 100% of the event was not warned (see the last paragraph in section 4.2.1), the maximum event lead time may not be representative of the entire area of the event. In these situations, an ***area weighted lead time*** is calculated to correct for the portion of the flooded area that was not warned. For example, the percentage of an event warned was 83, and the lead time across the warned portion of the flooded area was 60 minutes. To properly account for the 17% of the event not warned, the 60-minute lead time is multiplied by 83%, resulting in an area weighted lead time equal to 49.8 minutes.

4.2.4 Backup Mode for Warnings. All warnings issued by the backup office are attributed to the primary WFO, listed in the WMO header of the warning.

4.3 RFC River Stage Forecasts. The RFCs operate the river stage forecast verification software, and the OCWWS Hydrological Services Division maintains policy. For a selected set of locations, both stream level observations (stage) and stage forecasts issued by RFCs are posted to a verification database at each RFC. Forecast values are matched with concurrent observations. From these pairs, verification statistics measuring the performance of the forecast system are calculated. The initial phase of river forecast verification is based on calculations of mean, mean absolute, and root mean square differences between observed and forecast values for each verification site on the river.

No later than the 20th of each month, verification statistics for the previous month's data are sent from the RFCs to the OCWWS Performance Branch. NWS employees access verification statistics through the *Stats on Demand* feature of the NWS Performance Management website. *Stats on Demand* uses an interactive database to generate verification statistics customized to the user's request. The system allows verification statistics for locations to be grouped together by forecast lead time as well as hydrologic characteristics, i.e., (1) locations responding rapidly to rainfall, (2) locations with intermediate responses, and (3) locations with slow responses.

5. Quantitative Precipitation Forecast (QPF). Quantitative precipitation forecast verification statistics for the CONUS are found on the National Precipitation Verification Unit (NPVU) Web Page, which is operated and maintained by the OCWWS Performance Branch.

5.1 Data. Forecast, observation, and guidance data are collected and stored at the NPVU, where the verification statistics are computed and displayed on the hydrology verification page of the NWS Performance Management website.

The forecast data come from four sources:

a. The Environmental Modeling Center (EMC) runs the model guidance.

b. The Hydrometeorological Prediction Center (HPC) issues 10-km gridded guidance forecasts for the CONUS. These forecasts are prepared by forecasters who specialize in QPF.

c. The twelve CONUS RFCs collaborate with the WFOs in their respective forecast areas to prepare 10-km gridded QPFs. These forecasts are incorporated into the NWS River Forecast System.

d. The 116 CONUS WFOs each focus on their individual forecast areas and collaborate with the appropriate RFCs to prepare the gridded QPFs that are one of the forecast elements in the 5-km NDFD.

The quantitative precipitation estimate (QPE) product is the observation analysis used to verify all forecasts and guidance. This multi-sensor product, prepared by each CONUS RFC, uses rain gage, radar, and satellite data and is issued on the 4-km Hydrologic Rainfall Analysis Project

(HRAP) grid. The NPVU takes the QPE from each CONUS RFC and mosaics them into a national 4-km QPE. The Verification process compares each QPF to its time- and space - appropriate QPE, measures the forecast error, and calculates statistics that help assess forecast quality. These verification statistics are computed and displayed as two separate systems. The RFC forecasts are compared to all stored model and HPC guidance products (section 5.2), and the WFO forecasts are compared to all stored model and HPC guidance products (section 5.3).

5.2 Verification of RFC-issued QPFs. On a daily basis, each RFC forwards the four 6-hour periods of multi-sensor QPE on the 4-km HRAP grid to the NPVU, starting with the previous day's 6-hour QPE ending at 1800 UTC and ending with the current day's 6-hour QPE ending at 1200 UTC. Monthly, the QPEs, the RFC QPFs, the HPC QPFs, and model QPFs are re-mapped to a 32-km grid and used to compute 32-km verification statistics for each CONUS RFC forecast area and the entire CONUS. Monthly, a similar remapping process is also performed to the 4-km HRAP grid to compute 4-km verification statistics. Both resolutions of these verification statistics are computed for each month, each cold season (October to March), each warm season (April to September), each fiscal year, and each calendar year.

5.3 Verification of the WFO-issued NDFD QPFs. Monthly, the QPEs, the NDFD QPFs, the HPC QPFs, and the model QPFs are re-mapped to a 32-km grid and used to compute 32-km verification statistics for each CONUS WFO forecast area and the entire CONUS. Monthly, a similar remapping process is also performed to the 4-km HRAP grid to compute 4-km verification statistics. Both resolutions of these verification statistics are computed for each month, each cold season (October to March), each warm season (April to September), each fiscal year, and each calendar year.

5.4 HPC QPF Verification. The HPC also computes verification statistics for its QPFs and corresponding model QPFs. These data have been calculated since 1971 and are posted to the HPC website.

6. Aviation Verification Procedures.

6.1 TAF Verification. This *Stats on Demand* program is the official NWS TAF verification tool. TAFs are evaluated every five minutes, twelve times per hour or 288 times for an entire 24-hour TAF. The 5-minute interval times end in either a “0” or “5”. Forecast conditions at the end of each 5-minute interval are matched with the most recently reported METAR/SPECI, and each element (e.g., ceiling) is verified separately. Routine hourly METARs that do not report just before the hour are assumed to be missing, and all 5-minute verification intervals following that scheduled METAR are discarded as missing until a new METAR or SPECI is reported.

6.1.1 Verification Sites. All terminals for which the NWS issues TAFs may be verified. An interactive directory of all TAF verification sites appears on the NWS Performance Management website.

6.1.2 Data Input. OCWWS automatically collects all data from operational products. Forecast data come from the TAFs and observation data come from the METAR/SPECIs. All METARs and SPECIs are tested for reliability and consistency, and suspicious data are removed. These quality assurance algorithms are found on the NWS Performance Management website. Guidance data come from the alphanumeric MOS products derived from the GFS model and NAM. The Local AWIPS MOS Program (LAMP) product, derived from the GFS and available for CONUS locations, is available as a guidance product. The latest version of guidance available at TAF issuance time is used. The persistence forecast, defined as the observed conditions at the beginning time of the TAF, is also available as a guidance product. Forecaster identification, when appropriate, is read from a separate AWIPS product transmitted by the WFO with the WMO header: NTXX98 Kccc, where ccc is the WFO forecast office identifier.

6.1.3 TAF Verification Reports. NWS employees access verification statistics through the *Stats on Demand* feature of the NWS Performance Management website. *Stats on Demand* uses an interactive database to generate verification statistics customized to the user's request. The user is able to request data for any TAF element, a single forecast type (e.g. prevailing, TEMPO) and, if desired, corresponding data from a single guidance product, (i.e., GFS MOS, NAM MOS, GFS LAMP, or persistence) for one or more:

a. Dates, defined by the beginning and ending dates (format *mm/dd/yyyy).* Data more than 18 months old are only available in terms of full months.

b. Scheduled TAF beginning times, i.e., 0000, 0600, 1200, 1800 UTC.

c. Projection period groups (see section 6.1.4).

d. Verification sites (single site, multiple sites, WFO forecast area, states, NWS regions). When a single WFO or a subset of a WFO is selected, each forecaster has the option of requesting verification statistics that include only the TAFs issued by that forecaster. Similarly, each supervisor (usually the MIC or someone designated by the MIC) has the option of requesting verification statistics that include only the TAFs issued by forecasters he/she supervises. This privacy is accomplished through a system of usernames and passwords.

The user of *Stats on Demand* also specifies one of the following options concerning scheduled and amended TAFs: (a) verify scheduled TAFs only, (b) verify amended TAFs only, or (c) verify scheduled and amended TAFs.

Most verification statistics are computed from categorical contingency tables of forecasts versus observations for TAFs and the user-selected guidance product. Since forecasts are evaluated every 5 minutes, the contingency tables usually contain twelve entries per hour per verification site. However, the user of *Stats on Demand* can request that contingency table entries be given as the number of 5-minute intervals (frequency), number of hours, number of minutes, or percent of the contingency table total. Forecast categories for each element are defined in section 6.1.5.

6.1.4 Projections. Scheduled TAFs are issued and verified for projections of 24 hours beyond the initial valid time of the most recent scheduled TAF. For verification purposes, projections are defined from the initial valid time of the TAF, which is 0000, 0600, 1200, or 1800 UTC for scheduled TAFs and the issuance time for amendments.

a. When the user requests verification statistics for scheduled TAFs only, he/she may select one or more of the following projection period groupings:

- Greater than zero to 3 hours.
- Greater than 3 to 6 hours.
- Greater than 3 to 6 hours.
- Greater than 6 to 9 hours.
- Greater than 9 to 12 hours.
- Greater than 12 to 18 hours.
- Greater than 18 to 24 hours.

b. When the user requests verification statistics for amended TAFs only or scheduled and amended TAFs combined, he/she selects one or both of the following projection periods:

- Greater than zero to 3 hours.
- Greater than 3 to 6 hours.

6.1.5 Elements. The user of *Stats on Demand* specifies a single element. To receive results for multiple elements, the user must run *Stats on Demand* separately for each element desired.

a. Ceiling Height.

(1) Ceiling height is recorded in the database in hundreds of feet AGL and verified in the following categories. From these categories, contingency tables of forecasts versus observations and guidance versus observations are prepared, and verification statistics are computed, e.g., ESS, Peirce Skill Score (PSS), and percent hits. Sometimes categories are combined in computing the ESS due to zero or low population of the rare event categories. POD, FAR, and CSI are also computed for various categories.

- Less than 200 feet.

- 200 to 400 feet.
- 500 to 900 feet.
- 1000 to 1900 feet.
- 2000 to 3000 feet.
- Greater than 3000 feet; including cases with no ceiling.

(2) A 2-category verification is also available whenever the user selects a critical threshold value, *x* (in hundreds of feet).

- Ceiling less than *x*.
- Ceiling greater than or equal to *x* (includes cases with no ceiling).

b. Visibility. Visibility is recorded in the database in statute miles and fractions thereof and verified in the following categories. For observations taken on or after October 19, 2007, the surface visibility is always used. Whenever the surface visibility is lower than the tower visibility, the surface visibility appears in the remarks section of the observation. However, for observations taken before October 19, 2007, the visibility reported in the main body of the observation is used, regardless of whether it was a surface or tower visibility.

(1) From the following categories, contingency tables of forecasts versus observations and guidance versus observations are prepared, and verification statistics are computed, e.g., ESS, PSS, and percent hits. Sometimes categories are combined in computing the ESS due to zero or low population of the rare event categories. POD, FAR, and CSI are also computed for various categories.

- Less than ½ statute mile.
- ½ to less than 1 statute mile.
- 1 to less than 2 statute miles.
- 2 to less than 3 statute miles.
- 3 to 5 statute miles.
- Greater than 5 statute miles.

(2) A 2-category verification is also available whenever the user selects a critical threshold value, *y* (in statute miles).

- Visibility less than *y*.
- Visibility greater than or equal to *y*.

c. Flight Category. To determine the flight category, the ceiling and visibility are each converted to the categories in Table 8. The categories for ceiling and visibility are then combined by taking the lower category of the two. This is the flight category. Contingency tables of forecasts versus observations and guidance versus observations are prepared, and verification statistics are computed, e.g., ESS, PSS, and percent hits. Sometimes categories are combined in computing the ESS due to zero or low population of the rare event categories. POD, FAR, and CSI are also computed for various categories.

A 2-category verification is also available whenever the user selects the following critical threshold values: *x* (in hundreds of feet) for ceiling and *y* (in statute miles) for visibility, which are defined by the user.

(1) Ceiling less than *x* or visibility less than *y*.

(2) Ceiling greater than or equal to *x* and visibility greater than or equal to *y*.

Table 8. Categories for ceiling and visibility used to determine the flight category.

CATEGORY	CEILING (feet)	VISIBILITY (statute miles)
Very Low Instrument Flight Rules (VLIFR)	less than 200	less than ½
Low Instrument Flight Rules (LIFR)	200 to 400	½ to less than 1
Instrument Flight Rules (IFR)	500 to 900	1 to less than 3
Marginal Visual Flight Rules (MVFR)	1000 to 3000	3 to 5
Visual Flight Rules (VFR)	no ceiling or greater than 3000	greater than 5

d. Wind Direction. From the following categories, contingency tables of forecasts versus observations and guidance versus observations are prepared, and verification statistics are computed, e.g., ESS, PSS, and percent hits. Wind direction is not verified whenever the observed speed is less than 6 knots or the observed or forecast direction is unspecified due to calm or variable winds.

(1) North (340 to 20 degrees).

(2) Northeast (30 to 60 degrees).

(3) East (70 to 110 degrees).

(4) Southeast (120 to 150 degrees).

(5) South (160 to 200 degrees).

(6) Southwest (210 to 240 degrees).

(7) West (250 to 290 degrees).

(8) Northwest (300 to 330 degrees).

e. Sustained Wind Speed. From these categories, contingency tables of forecasts versus observations and guidance versus observations are prepared, and verification statistics are computed.

(1) Less than 8 knots.

(2) 8 to 12 knots.

(3) 13 to 17 knots.

(4) 18 to 22 knots.

(5) 23 to 27 knots.

(6) 28 to 32 knots.

(7) Greater than 32 knots.

f. Wind Gusts. From these categories, contingency tables of forecasts versus observations are prepared, and verification statistics are computed. MOS guidance is not available for wind gusts.

(1) No gusts or gusts less than 18 knots.

(2) 18 to 22 knots.

(3) 23 to 27 knots.

(4) 28 to 32 knots.

(5) 33 to 37 knots.

(6) 38 to 42 knots.

(7) 43 to 47 knots.

(8) Greater than 47 knots.

g. Significant Weather Type. Each of the following significant weather types is verified separately in 2-category contingency tables of forecasts versus observations and guidance versus observations. The two categories comprising each of these contingency tables are occurrence and non-occurrence of the weather type. Precipitation intensity is not verified. Note: To get the most complete set of scores, this element should be verified *without* guidance since all guidance products issue these forecasts for a very limited number of weather types. GFS MOS and NAM MOS only forecast significant weather type (5). GFS LAMP only forecasts the following significant weather types as deterministic, binary variables: (1), (2), (4), and (5) and (6).

(1) Liquid precipitation—rain (RA), rain showers (SHRA), drizzle (DZ).

(2) Snow types—snow (SN), snow showers (SHSN), snow grains (SG).

(3) Ice types, i.e., ice crystals (IC), ice pellets (PL), showers of ice pellets (SHPL), small (less than 1/4 inch diameter) hail/snow pellets (GS), showers of GS (SHGS).

(4) Freezing precipitation—freezing rain (FZRA), freezing drizzle (FZDZ).

(5) Fog/Mist—Fog (FG), mist (BR), and freezing fog (FZFG).

(6) Haze (HZ) and smoke (FU).

(7) Thunderstorms (TS), including funnel clouds (FC) and tornadoes/waterspout (+FC). Some observation stations do not report thunderstorms. These METARs use the TSNO remark. Thunderstorms are not verified whenever the TSNO remark appears in the observation. VCTS are not considered in verification and are ignored whenever they appear in forecasts or observations. Note: Vicinity Thunderstorms (VCTS) means thunderstorms are forecast or were observed within a 5- to 10-mile radius from the center of a terminal.

(8) Hail (1/4 inch or greater diameter) (GR) and showers of GR (SHGR).

(9) Squalls (SQ).

(10) Blowing snow (BLSN), drifting snow (DRSN).

(11) Blowing spray (BLPY).

(12) Volcanic ash (VA).

(13) All dust and sand events, i.e., widespread dust (DU), blowing dust (BLDU), drifting dust (DRDU), dust storm (DS), sand/dust whirls (PO), blowing sand (BLSA), drifting sand (DRSA), and sandstorm (SS).

6.1.6 Forecast Types. TAFs primarily predict prevailing conditions and use the “from” (FM) change indicator to introduce changes to the forecast prevailing conditions. Prevailing forecast verification is described in section 6.1.6.1. Another “type” of forecast, called the operational impact forecast, is defined in section 6.1.6.2. Sometimes a TEMPO or PROB change indicator is used to respectively designate a temporarily fluctuating or probabilistic forecast condition. When a TEMPO or PROB change indicator is used, two forecasts are valid for the same time. TEMPO and PROB forecast evaluation are explained, respectively, in sections 6.1.6.3 and 6.1.6.4. The following terms will be repeated several times in sections 6.1.6.1 through 6.1.6.4 are defined:

a. Change.

(1) For ceiling and visibility, change is defined as at least a one-category change. The categories for each of these elements are defined in section 6.1.5, respective paragraphs a. and b.

(2) Each of the thirteen significant weather types is a binary variable, and change is defined as the starting or stopping of that weather type. Precipitation intensities are ignored.

(3) For wind direction, change is defined

- as a 40-degree or greater wind shift between successive observations, considering only 6-knot or greater sustained wind speeds, or

- by the presence of a variable wind remark.

(4) For sustained wind speed, change is defined as at least an 8-knot increase or decrease between successive observations.

(5) For wind gusts, change is defined

- as at least a 10-knot increase or decrease between successive observations, or
- when successive observations change from the existence of gusts to the existence of no gusts or vice versa.

b. Hit.

(1) For ceiling and visibility, a forecast hit is defined as the forecast category equaling the observation category. Categories are defined in section 6.1.5, respective paragraphs a. and b.

(2) For each of the thirteen significant weather types, a hit occurs when the forecast and observation agree on the occurrence of that weather type.

(3) For wind direction, a hit occurs whenever the absolute error is 30 degrees or less.

(4) For sustained wind speed, a hit occurs whenever the absolute error is less than 8 knots.

(5) For wind gusts, a hit occurs whenever the absolute error is less than 10 knots, or neither observation nor forecast contains gusts. Forecast and observed gusts less than 16 knots are treated as *no gusts*.

c. Less [More] in Error. When comparing two forecast types (i.e., prevailing and TEMPO, prevailing and PROB):

(1) For ceiling or visibility, less [more] in error means the TEMPO or PROB forecast was not a hit and had a smaller [larger] absolute categorical error than the prevailing forecast. Use the categories defined for ceiling and visibility in section 6.1.5, respective paragraphs a. and b.

(2) For wind direction, sustained wind speed, and wind gusts, less [more] in error means the TEMPO or PROB forecast was not a hit, and the absolute error of the TEMPO or PROB forecast was lower [higher] than the absolute error of the prevailing forecast.

(3) All thirteen significant weather types are binary variables, so the term "less [more] in error" is not used when referring to any of them.

d. More [Less] Favorable Flight Conditions. When comparing two forecast types (i.e., prevailing and TEMPO, prevailing and PROB):

(1) For ceiling or visibility, the more [less] favorable flight conditions are defined as the higher [lower] category forecast, using the categories defined for each element in section 6.1.5, respective paragraphs a. and b.

(2) For each of the thirteen significant weather type forecasts (each is a binary variable), the more [less] favorable flight conditions are defined as the negative [positive] forecast of the event.

(3) For sustained wind speed and wind gust forecasts, the more [less] favorable flight conditions are defined as the lower [higher] speed forecast. "No gust" forecasts are more favorable than gust forecasts and vice versa.

(4) Wind direction forecast are not compared in this manner.

6.1.6.1 Prevailing Forecast. The prevailing forecast is defined as (1) the forecast conditions that are in the initial time period of the TAF and (2) any forecast conditions that immediately follow a FM change indicator. For the element specified by the user of *Stats on Demand* (e.g., ceiling), the prevailing forecast is evaluated at the end of every 5-minute interval of the TAF by comparing it to the most recent METAR/SPECI available. Most verification is categorical, using the categories defined in section 6.1.5, and results are recorded twelve times per hour in contingency tables of forecasts versus observations. Prevailing forecasts may be evaluated by themselves, or they may be matched with one guidance product at a time, producing an additional contingency table of guidance forecasts versus observations. Conventional verification statistics are computed from the contingency tables, and comparisons may be drawn between prevailing forecast and guidance performance.

6.1.6.2 Operational Impact Forecast (OIF). TAFs are sometimes formatted in a manner whereby two forecasts are valid for a single terminal at the same time. One of the following circumstances applies to all NWS TAFs at all times: (1) Just the prevailing forecast is in effect, (2) The prevailing forecast is in effect simultaneously with a forecast for temporary conditions (TEMPO), or (3) The prevailing forecast is in effect simultaneously with a 30% probabilistic forecast (PROB). For verification, the OIF is defined as the forecast in effect that is most likely to have the largest impact on operations. The following rules are used to determine the OIF:

a. The OIF is undefined for wind direction.

b. If no TEMPO or PROB forecast is in effect for the user-specified element, then the OIF for that element is defined as the prevailing forecast.

c. If a PROB forecast is in effect for the user-specified element, then the OIF for that element is defined as the forecast (prevailing or PROB) of the less favorable flight conditions, i.e., lower ceiling category, lower visibility category, higher wind speed, or the occurrence of the weather type.

d. If a TEMPO forecast is in effect for the user-specified element, then the OIF for that element is defined through a two step process.

(1) *First step—the variability test.* The legitimacy of the TEMPO forecast is first evaluated by a variability test that is performed at the end of every 5-minute interval of the TAF TEMPO forecast. If the observation database changes twice or more during the variability period, then that 5-minute interval of the TEMPO forecast passes the variability test. The term "change" for each forecast element in the TAF is defined at the beginning of section 6.1.6. The beginning time of the variability period for each 5-minute window in the TEMPO forecast is defined as 90 minutes prior to the ending time of the 5-minute interval being tested. The end time of the variability period for each 5-minute window in the TEMPO forecast is defined as 90 minutes after the ending time of the 5-minute interval being tested. Note: This test only measures condition variability—it does not measure forecast correctness.

(2) *Second step.* If the TEMPO forecast fails the variability test for a given 5-minute interval, then the OIF for that interval is defined as the forecast with the less favorable flight conditions, i.e., lower ceiling category, lower visibility category, higher wind speed, or the occurrence of the significant weather type.

If the TEMPO forecast passes the variability test for a given 5-minute interval, then the OIF for that interval is (1) defined as the forecast with the smallest categorical error for ceiling and visibility (in a tie, the OIF is set equal to the prevailing forecast category); (2) defined as the forecast with the smallest absolute error for wind speed and wind gusts (in a tie, the OIF is set equal to the prevailing forecast); or (3) set equal to the observation (no error) for each significant weather type.

The OIF for flight category is determined by first calculating the OIF separately for ceiling and visibility. Then, the OIFs for ceiling and visibility are each converted to the categories in Table 8. The lower category of the two is the flight category OIF.

Just like the prevailing forecast, the OIF is evaluated only for the element specified by the user of *Stats on Demand* at the end of every 5-minute interval that the TAF is valid. At each of these times, the OIF is compared to the most recent METAR/SPECI available. Most verification is categorical, using the categories defined in section 6.1.5, and results are recorded twelve times per hour in contingency tables of forecasts versus observations. OIFs may be evaluated by themselves, or they may be matched with one guidance product at a time, producing an additional contingency table of guidance forecasts versus observations. Conventional verification statistics are computed from the contingency tables, and comparisons may be drawn between OIF performance and guidance performance.

6.1.6.3 TEMPO Forecast. The TEMPO forecast is evaluated at the end of every 5-minute interval that the TEMPO forecast is valid for the user-specified element. TEMPO forecast evaluation is separate from OIF evaluation, but some of the same methodology is employed for TEMPO evaluation as OIF evaluation. Since no guidance product provides TEMPO forecasts, TEMPO forecast verification statistics are not matched with guidance. The following statistics are tallied:

a. Number of hours each significant weather type was observed. This is the number of hours each significant weather type was observed regardless of whether or not it was included in the prevailing, TEMPO or PROB30 forecast. Note: By definition, this statistic is only valid for the significant weather types and is not provided for the other elements.

b. Number of hours TEMPO forecast. This is the total number of hours (the number of 5-minute intervals divided by 12) that TEMPO forecasts were issued for the user-specified element. Data are given to the nearest hour.

c. Justified TEMPO. This is the total number of hours when a TEMPO forecast containing the user-specified element passed the OIF variability test and was, therefore, justified. The OIF variability test is described in section 6.1.6.2, paragraph d.(1). This statistic only measures condition variability—it does not measure forecast correctness. *Example: A TEMPO group is in effect from 0800 until 1200 UTC. The end of every 5-minute interval must be checked for justification. Start with the end time of 0800-0805 UTC and see if two or more changes occur between 0635 and 0935 UTC (0805 UTC ±90 minutes). If a 1500-foot ceiling at 0635 UTC rises to 2500 feet at 0720 UTC, and then drops to 1200 feet at 0840 UTC, then two changes occurred between 0635 and 0935 UTC, making the TEMPO group justified for the 0800-0805 UTC interval. Repeat this process for every five minute interval until you finish the TEMPO group at noon (last 5-minute interval is 1155-1200 UTC). Assuming no more ceiling category changes occurred after 0840 UTC, the number of 5-minute intervals with a justified TEMPO forecast was 10, which converts to 0.8 (10/12) hour. Ideally, this number should equal the total number of TEMPO forecast hours, which in this example was 4.0.*

d. Justified TEMPO–Hit (%). Considering only the 5-minute intervals when the TEMPO forecast was justified for the user-specified element, this is the percentage of time that the TEMPO forecast was a hit. Ideally, this statistic ranges between 10 and 49. *Example: Between 0600 and 0820 UTC, the* *observations indicated that ceilings varied sufficiently to justify a TEMPO group. The TAF prevailing group forecast ceilings at 800 feet, the TEMPO group forecast ceilings at 300 feet, and ceilings 200 to 400 feet, inclusive, were observed at the end of 40% of the 5-minute intervals between 0600 and 0820 UTC. Justified TEMPO - Hit (%): 40.*

e. Justified TEMPO–Improved the TAF (%). Considering only the 5-minute intervals when the TEMPO forecast was justified, this is the percentage of time the TEMPO forecast was less in error than the prevailing forecast, without being a hit. Since each of the thirteen significant weather types is a binary variable and each can only hit or miss, they are not evaluated with this statistic. Ideally, 10 to 49 percent of TEMPO forecasts are hits (previous statistic), and this statistic is zero. *Example: Between 0600 and 0820 UTC, the observations indicated that ceilings varied enough to justify a TEMPO group. The TAF prevailing group forecast ceilings at 1200 feet, the TEMPO group forecast ceilings at 700 feet, and ceilings between 200 and 400 feet were observed at the end of 40% of the 5-minute intervals between 0600 and 0820 UTC. Justified TEMPO -Improved TAF (%): 40.*

f. Unjustified TEMPO (hours). This is the total number of hours that the TEMPO forecast failed the OIF variability test and was, therefore, unjustified. It is determined by subtracting the number of hours of justified TEMPO forecasts subtracted from the total number of TEMPO forecast hours for each user-selected element. The OIF variability test is described in section 6.1.6.2, paragraph d.(1).

g. Unjustified TEMPO - Should Be FM (%). Considering only the 5-minute intervals when the TEMPO forecast was not justified, this statistic is the percentage of time when the TEMPO forecast was a hit, resulting in an incorrect prevailing forecast. Ideally, this statistic is zero. *Example: During the period that the observations indicated that ceilings did not vary enough to justify a TEMPO group, the TAF prevailing group forecast ceilings at 1200 feet, the TEMPO group forecast ceilings at 800 feet, and ceilings were observed between 500 and 900 feet all the time. TEMPO S/B FM (%): 100.*

h. Unjustified TEMPO - Benign (%). Considering only the 5-minute intervals when the TEMPO forecast was not justified, this statistic is the percentage of time whenever (a) the TEMPO forecast was more in error than the prevailing forecast, and (b) the TEMPO forecast predicted *more* favorable flight conditions than the prevailing forecast. In these cases, poor TEMPO forecasts are benign to flight operations because the pilot has already planned for the less favorable flight conditions in the prevailing forecast. Wind direction is not evaluated with this statistic. Ideally, this statistic is zero. *Example: The TAF prevailing group forecast ceilings at 700 feet, the TEMPO group forecast ceilings at 1200 feet, and ceilings were observed between 500 and 900 feet at the end of 90% of the 5-*

minute intervals that failed the justification test. Tempo Benign (%): 90.

i. Unjustified TEMPO - Hurt (%). Considering only the 5-minute intervals when the TEMPO forecast was not justified, this statistic is the percentage of time whenever (a) the TEMPO forecast was more in error than the prevailing forecast, and (b) the TEMPO forecast predicted *less* favorable flight conditions than the prevailing forecast. In these cases, poor TEMPO forecasts hurt flight operations because the pilot is forced to plan for the less favorable flight conditions that did not occur. Wind direction is not evaluated with this statistic. Ideally, this statistic is zero. *Example: The TAF prevailing group forecast ceilings at 1400 feet, the TEMPO group forecast ceilings at 600 feet, and ceilings were observed between 1000 and 1900 feet at the end of 90% of the 5-minute intervals that failed the justification test. TEMPO Hurt (%): 90.*

Most of the nine aforementioned statistics are used for every element, including the elements collectively called "significant weather type." However, the following exceptions exist.

Each significant weather type (e.g., rain, fog, snow) is a binary variable, i.e., it occurs or it doesn't (yes or no). Therefore, statistic (5), "improved the TAF" has no meaning for any of the significant weather types because a binary forecast situation that is a "miss" cannot simultaneously improve the forecast, i.e., the hits have already been counted by statistic (d).

Statistic (a) is only used for the significant weather types because it lists the total number of hours that each significant weather type occurred and has no meaning for elements such as ceiling, visibility, flight category and wind. All of the significant weather types collectively fit into the format of Table 9 in the *Stats on Demand* reports:

Table 9. Example of TEMPO table for the significant weather types.

TEMPO FORECASTS (EVALUATION BY 5-MINUTE INTERVALS)

	Condition Observed (Hours)	TEMPO Forecast (Hours)	Justified TEMPO		Unjustified TEMPO			
			Number of Hours	Hits % of (d)	Number of Hours	TEMPO S/B FM %of (f)	TEMPO Benign % of (f)	TEMPO Hurt % of (f)
(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)
LIQUID	1296	1305	533.7	42%	771.3	37%	1%	55%
etc.								

6.1.6.4 PROB Forecast. The PROB forecast is evaluated at the end of every 5-minute interval for the user-specified element. Since no guidance product provides PROB forecasts, PROB forecast verification statistics are not matched with guidance. The following statistics are tallied:

a. Number of Hours: This is the total number of hours (the number of 5-minute intervals divided by twelve) that PROB groups were valid for the user-specified element. Data are given to the nearest hour.

b. PROB Hit (Element + precip/TS) (%): This is the percentage of all 5-minute intervals within PROB groups for the user-specified element that (a) were forecast hits and (b) precipitation or a thunderstorm occurred. Credit is not granted if the user-specified element is a hit, but precipitation or a thunderstorm did not occur. All elements are eligible for evaluation except precipitation and thunderstorms. Ideally, this statistic is between 30 and 40. *Example: The prevailing forecast is 4000 feet, the PROB forecast is 1500 feet, and light rain is forecast with the lower ceilings. Ceilings between 1000 and 1900 feet with light snow were observed at the end of 30% of the 5-minute intervals. Prob Hit w/ precip/TS: 30. Note: The 30% hit rate occurred even though rain was forecast with the lower ceilings and snow was observed. For this statistic, any type of precipitation or a thunderstorm to verify the ceiling. The significant weather type (incorrect rain forecast) is verified separately in the significant WX type rows. If no precipitation had occurred with the lower ceilings, the forecaster would not have gotten credit for the ceilings and the Prob Hit w/ precip/TS would have been zero.*

c. PROB Hit w/out precip/TS (%): This is the percentage of 5-minute intervals that the user-specified element forecast in PROB groups was a hit, even though precipitation or a thunderstorm type (TS, FC, +FC) defined in previous bullet) did not occur. All elements are verified except for the following significant weather types: all precipitation types (rain types, snow types, ice types, freezing precipitation, hail) and thunderstorm types (TS, FC, +FC). For all precipitation types and thunderstorm types, this column is "blacked out." *Example: The prevailing ceiling forecast is 4000 feet, the PROB forecast is 1500 feet, and light rain is forecast with the lower ceilings. Ceilings between 1000 and 1900 feet were observed at the end of 30% of the 5-minute intervals, but no precipitation or thunderstorm events occurred at the end of these 5-minute intervals. Prob Hit w/out precip/TS: 30.*

d. PROB Hit (Precip/TS only) (%): This is the percentage of 5-minute intervals within PROB groups that were forecast hits. Only precipitation and thunderstorms are eligible for evaluation. Ideally, this statistic is between 30 and 40.

e. PROB Improved the TAF (%): This is the percentage of 5-minute intervals within PROB groups for the user-specified element whenever the PROB forecast was not a hit, but the PROB forecast was less in error than the prevailing forecast. Unlike "PROB Hit," credit is granted whenever precipitation or a thunderstorm does not occur with the user-specified element. All elements are eligible for evaluation except the thirteen weather types. Ideally, this statistic is zero. *Example: The TAF prevailing group forecast ceilings at 1200 feet, the PROB group forecast ceilings at 700 feet, ceilings below 200 to 400 feet were observed 40% of the time,* *and ceilings 1000 feet or higher were observed 60% of the time. Prob Imp (%): 40.*

f. PROB Benign (%). This is the percentage of all 5-minute intervals within PROB groups for the specified user-element whenever (a) the PROB forecast was more in error than the prevailing forecast, and (b) the PROB forecast predicted *more* favorable flight conditions than the prevailing forecast. In these cases, the poor direction is not eligible for evaluation. Ideally, this statistic is zero. *Example: The TAF prevailing group forecast ceilings at 700 feet, the PROB group forecast ceilings at 1200 feet, and ceilings were observed between 500 and 900 feet at the end of 90% of the 5-minute intervals. Prob Benign (%): 90.*

g. PROB Hurt (%). This is the percentage of all 5-minute intervals within PROB groups for the user-specified element whenever (a) the PROB forecast was more in error than the prevailing forecast, and (b) the PROB forecast predicted *less* favorable flight conditions than the prevailing forecast. In these cases, the poor PROB forecasts hurt flight operations, because the pilot is forced to plan for the less favorable flight conditions that ultimately do not occur. No check is made to see if precipitation or thunderstorms occurred with the other elements and weather types. Wind direction is not evaluated with this statistic. Ideally, this statistic is zero. *Example: The TAF prevailing group forecast ceilings at 1400 feet, the PROB group forecast ceilings at 600 feet, and ceilings were observed between 1000 and 1900 feet 90% of the time. Prob Hurt (%): 90.*

6.2 Aviation Weather Center (AWC) Verification Procedures.

6.2.1 Background. The AWC uses the automated Real-Time Verification System (RTVS), created specifically for verifying AWC's manually produced forecasts and various associated automated forecast algorithms. RTVS is continuously under review and revision as more and better sources of aviation verification observations are implemented. Verification techniques are under constant scrutiny in an effort to improve upon the subjectivity of pilot reports and other observations/observation products used in many aviation forecast verification procedures. Additionally, the RTVS' convective verification procedures are often revised and refined in an effort to provide the AWC with the best possible statistics for describing the accuracy of its convective forecasts. The National Convective Weather Diagnostic algorithm is currently used to verify AWC's convective products. While RTVS provides a baseline and a starting point for verification trend monitoring, the statistics are subject to change as RTVS evolves into a more mature system meeting the AWC's needs. Statistics are also prone to substantial monthly and seasonal variability based on the subjectivity and unreliable frequency of pilot reports. No standardized observing network exists for verifying aviation forecast variables, such as icing and turbulence. Despite these problems, statistics are presented as 12-month running averages.

6.2.2 Domestic Products Verified and Statistics Calculated.

a. Airman's Meteorological Information (AIRMET).

(1) Icing (AIRMET Zulu) and Turbulence (AIRMET Tango). The following verification statistics, defined in appendix A section 4.4, are calculated separately for AIRMET Zulu and AIRMET Tango: POD, POD of no observations (POD[N]), the percent area of AIRMET coverage across the domestic airspace (% Area), and the percent volume of AIRMET coverage across the domestic airspace.

(2) Instrument Flight Rules (IFR) Conditions (AIRMET Sierra). The following verification statistics are calculated: POD, FAR, and % Area.

b. Convective Forecasts.

(1) Convective Significant Meteorological Information (SIGMET). The following verification statistics are calculated: POD, FAR, % Area.

(2) Collaborative Convective Forecast Product: The following verification statistics are calculated: POD, FAR, and % Area.

7. Tropical Cyclone Verification Procedures. The National Hurricane Center (NHC) and the Central Pacific Hurricane Center (CPHC) verify tropical cyclone track and intensity forecasts.

7.1 Tropical Cyclone Forecasts/Advisories. NHC and CPHC issue Tropical Cyclone Forecast/Advisory products. The Tropical Cyclone Forecast/Advisory product will be referred to as the TCM product in this instruction. The first TCM product associated with each tropical system is normally issued when meteorological data indicate the formation of a tropical or subtropical cyclone. Subsequent advisories are issued at 0300, 0900, 1500, and 2100 UTC. Special forecasts/advisories are issued if significant changes to the forecast occur. Each advisory product contains 12-, 24-, 36-, 48-, 72-, 96-, and 120-hour forecast positions and maximum sustained wind speed. Forecast positions are rounded to the nearest tenth of a degree of latitude and longitude, and forecast intensities are rounded to the nearest 5 knots.

7.1.1 Verification Elements. The following TCM elements are verified at 12, 24, 36, 48, 72, 96, and 120 hours:

a. Maximum Sustained Surface Wind. A tropical cyclone's intensity is verified by the maximum sustained surface wind, defined as the highest 1-minute average wind (at an elevation of 10 m with an unobstructed exposure) associated with the cyclone at a particular point in time. Units for this element are "knots."

b. Location. The position of the tropical cyclone center is usually defined by the cyclone's minimum wind or minimum pressure at the surface. Units for this element are "degrees latitude and longitude."

7.1.2 Verification Process. Each TCM product contains an operational estimate of the tropical cyclone's current location and maximum sustained surface wind speed. These estimates are determined from a variety of sources, including surface observations from land or marine platforms, aircraft reconnaissance data, radars, and satellites. During a tropical cyclone event as new observations become available, an ongoing evaluation of the operational location and intensity estimates results in the creation of a "working best track", whose points will often differ from the operational values contained in the TCM. A preliminary verification of the TCM forecast parameters can be accomplished by comparison with the working best track.

After each tropical cyclone event has concluded, hurricane specialists review all available data and refine the working best track. The refined set of locations and intensities is known as the "final best track". A cyclone's final verification is performed by comparing the TCM location and intensity forecasts with the final best track. In order to be included in the verification sample, the system must have been a tropical (or subtropical) cyclone at both the initial time and the forecast time. A second verification is often conducted in which the depression stage is omitted.

Preparation of a cyclone's final best track is a time-consuming process that may not be completed until several weeks after the conclusion of the event. As a result, final verifications for the season are generally not available at the conclusion of the hurricane season.

7.2 Model Verification. A variety of models are run operationally and provide forecasted tropical cyclone tracks. Several models provide forecasted tropical cyclone intensities. The models range in complexity from simple statistical models to three-dimensional primitive equation models.

7.2.1 Verification Elements. The following model elements may be verified at 12, 24, 36, 48, 72, 96, and 120 hours:

a. Maximum Sustained Surface Wind. A tropical cyclone's intensity is verified by the maximum sustained surface wind, defined as the highest 1-minute average wind (at an elevation of 10 m with an unobstructed exposure) associated with the cyclone at a particular point in time. Units for this element are "knots."

b. Location. The position of the tropical cyclone center, usually defined by the cyclone's minimum wind or minimum pressure at the surface. Units for this element are "degrees latitude and longitude."

7.2.2 Verification Process. A preliminary verification of model location and intensity forecasts may be made against the working best track. The final verification will be made using the final best track.

7.3 Verification Reports. The NHC and the CPHC maintain verification statistics and post them on their respective Websites:

http://www.nhc.noaa.gov/verification
http://www.prh.noaa.gov/cphc/pages/hurrclimate.php

8. Climate Verification Procedures.

8.1 Medium Range and Seasonal Outlooks. The Climate Prediction Center (CPC) verifies its medium range and seasonal outlooks.

a. The following mean temperature and total precipitation forecasts are verified on a grid that covers the contiguous United States:

(1) 6 to 10 day forecast.

(2) Week 2 (8-14 day) forecast.

(3) Monthly, issued monthly for the following month with a 0.5-month lead.

(4) Seasonal, issued monthly for twelve consecutive 3-month seasons. Each of the twelve seasonal forecasts is issued with a 0.5-month through 12.5-month lead time.

b. The data specifications follow:

(1) Data Source: River Forecast Centers – Approximately 5000 stations per day are used, including approximately 1500 stations per day from the Hydrologic Automated Data System (HADS) and several hundred stations per day from the Climate Anomaly Data Base.

(2) Resolution: The station data are fit to a 0.5x0.5-degree grid, and the verification is done on a 2x2-degree grid.

(3) Domain: 20 to 60 degrees North; 60 to 140 degrees West.

(4) Format: The format is sequential 32-bit IEEE floating point created on a big endian platform (e.g. cray, sun, sgi and hp). The undefined (missing) value is 9999.

(5) Window: The Day 1 analysis is valid for the window from 1200 UTC on Day 0 (the day issued) to 1200 UTC on Day 1. Because of report receipt

timing, daily minima are available 1 day earlier than the daily maxima and the daily means.

(6) Analysis Scheme: Modified Cressman (1959) scheme (Glahn et al. 1985; Charba et al. 1992). The minimum number of stations required for analysis is 350. Whenever the number of stations is fewer than 350, the analysis is not performed for that particular day.

(7) Quality Control: A climatological standard deviation check is used. If a reported value is more than 4 standard deviations removed from the historical distribution, the value is omitted from the analysis.

A version of the Heidke Skill score (described in appendix A) is computed for verification.

8.2 U.S. Hazards Assessment Product. CPC verifies heavy precipitation forecasts in its 3- to 14-day U.S. Hazards Assessment Product. Hazard forecasts of daily (1200 to 1200 UTC) precipitation expected to exceed the hazard threshold at specific grid points on specific dates are made each weekday for the 3- to 14-day forecast period, e.g., a forecast made on Tuesday is valid 1200 UTC Friday (Day 3) until 1200 UTC on the Tuesday two weeks after the forecast is issued (Day 14). All issuances of the Hazard Assessment are verified. The forecast domain consists of a one-degree-latitude by one-degree-longitude grid (881 points) over the contiguous United States. The daily hazard threshold for each grid point is defined as the greater of one inch of precipitation for a given day or the 95th percentile of the climatology for a given day. For verification, daily (1200 to 1200 UTC) precipitation amounts are analyzed to each of the 881 grid points. One "event" is defined as any grid point where observed precipitation equals or exceeds the daily threshold.

A similar procedure is used for verifying severe weather hazards (tornadoes, damaging winds, and large hail) included in the hazard assessment product. Observation data are taken from the SPC preliminary severe weather reports.

The following 2x2 contingency table is used to classify all events and non-events with respect to how they were forecast:

Table 10. Special 2x2 contingency table.

		Forecasts	
		Yes	No
Events	Yes	A	B
	No	C	*X*

Any event that occurs on one or more days within the hazardous forecast area during the hazard period is counted as one "hit" (*A* in Table 10). For example, a heavy precipitation hazard was forecast for a particular grid point from November 17 thru 19. That grid point received enough precipitation to exceed its daily threshold on two separate dates: November 17 and 19. Consequently, one "hit" is counted. One "hit" is also counted whenever *no hazard* is forecast, and the observed precipitation does *not* equal or exceed the hazard threshold during any of the eleven forecast days *(X)*. A "miss" is counted whenever an event occurs with none forecast (*B*), or a hazard is forecast with no event reported (*C*; also known as a false alarm). From these counts, the following scores are computed (see appendix A, section 3): probability of detection, false alarm ratio, and threat score; the latter is also called the critical success index.

To compute the bias of the hazardous events (*A+B*), CPC uses the formula in Appendix A, section 3.4.

9. Model Verification Procedures. The Environmental Modeling Center verifies its numerical models. As part of its World Meteorological Organization responsibilities, the National Centers for Environmental Prediction Central Operations (NCO) sends monthly numerical model verification statistics to all World Forecast Centers. NCO also provides model verification statistics to the annual Numerical Weather Prediction report.

10. Use of Verification Information in Evaluating Forecaster Performance. Verification scores are not used to establish criteria for rating the forecasting and warning performance element of an individual's performance plan. Such use of the verification program is not appropriate because objectively derived verification scores by themselves seldom fully measure the quality of a set of forecasts. A forecaster demonstrates overall skill through his or her ability to analyze data, interpret guidance, and generate forecasts of maximum utility. Individual forecaster verification data is a private matter between office management and employees and will be safeguarded.

To properly utilize forecast verification scores in the performance evaluation process, managers use scores as an indicator of excellence or of need for improvement. For example, a skill score which is "clearly above average" may be used, in part, to recognize excellence via the awards system. However, NWS managers at all echelons should be aware that no two forecasters, offices, or management areas face the same series of forecast challenges. Factors which must be taken into account include the number of forecasts produced, availability and quality of guidance, local climatology, and the increased level of difficulty associated with rare events. There is no substitute for sound supervisory judgment in accounting for these influences.

11. References.

Charba J. P., A. W. Harrell III, and A. C. Lackner III, 1992: A monthly precipitation amount climatology derived from published atlas maps: Development of a digital database. NOAA TDL Office Note 92-7, 20 pp.

Cressman G. P., 1959: An operational objective analysis system. *Mon. Wea. Rev.*, **87**, 367–374.

Glahn H. R., T. L. Chambers, W. S. Richardson, and H. P. Perrotti, 1985: Objective map analysis for the local AFOS MOS Program. NOAA Tech. Memo. NWS TDL **75,** 34 pp.

Kluepfel, C.K., A.J. Schreiner, and D.A. Unger, 1994: The satellite-derived cloud cover product (sounder). NWS Technical Procedures Bulletin No. 410, NOAA, U.S. Department of Commerce, 15 pp.

Liu, W.T., K.B. Katsaros, and J.A. Businger, 1979: Bulk parameterization of air-sea exchanges of heat and water vapor including the molecular constraints at the interface. *J. Atmos. Sci.,* **36,** 1722-1735.

Weiss, S.J., D.L. Kelly, and J.T. Schaefer, 1980: New objective verification techniques at the National Severe Storms Forecast Center. Preprints, 8th Conference on Weather Forecasting and Analysis, Denver, Colorado, American Meteorological Society, 412-419.

APPENDIX A – Verification Scores

Table of Contents:

1. Introduction. Verification scores are applied at the local, regional, and national levels. Different scores may be applied to the same data. The type of score selected for use depends upon the objective. Frequently used scores are given in this manual and presented within the context of specific elements and events subject to verification. An excellent reference for verification scores is Wilks (1995).

In general terms, the scores are measures of accuracy and skill. **Accuracy** is a measure of how much a forecast agrees with the event or element being forecast. The smaller the difference between the forecast and observation, the greater the accuracy. **Skill** is a measure of improvement of a forecast over an established standard. Examples of standards often used for comparison include the climatological frequency (or value), persistence, or forecasts made by another process (e.g., model output statistics). The greater the improvement, the greater the skill.

2. Generalized Contingency Table. A generalized forecast/observation contingency table (Table A-1) is often used to summarize the forecast performance of a given element by category

(the term "category" is sometimes called class). The table is divided into k mutually exclusive and exhaustive categories. Each cell of the table, A_{ij} , gives the number of occurrences with the observation in the *ith* category (e.g., 13 to 17 knots for sustained wind speed) and the forecast in the *jth* category (e.g., 18 to 22 knots for sustained wind speed). Categorically correct forecasts (A_{ii} for all i), where all $i = j$, fall along the upper left to lower right diagonal of the contingency table. The row and column totals, R_i and C_i , respectively, are often called the marginal totals of the contingency table, and they are used in computing forecast bias and skill.

Table A-1. Generalized Contingency Table

	Forecast Category				
Observed Category	1	2	…	k	Total
1	A_{11}	A_{12}	...	A_{1k}	R_1
2	A_{21}	A_{22}	...	A_{2k}	R_2
…	...	...	...	...	...
k	A_{k1}	A_{k2}	...	A_{kk}	R_k
Total	C_1	C_2	...	C_k	N

The following scores may be computed from the data in this contingency table:

2.1 Percent Hits *(PH)* (also called percent correct) is the percentage of time categorical hits occurred *(i=j),* considering all categories. It is a measure of accuracy and may also be referred to as the categorical percentage correct.

$$PH = \frac{\sum_{i=1}^{k} A_{ii}}{N} x100$$

2.2 Bias by Category (BIAS) measures the tendency to over-forecast (*BIAS* greater than 1) or under-forecast (*BIAS* less than 1) a particular category, *i,* of a multi-category contingency table (see Table A-1, where *k* values of bias exist.

$$BIAS_i = \frac{C_i}{R_i}$$

2.3 Probability of Detection (POD). A *POD* may be calculated for each individual category, *i,* of Table A-1. It measures the forecaster's success in covering each event of category *i* with a correct forecast, A_{ii}. The *POD* does not penalize the forecaster for incorrect forecasts of category *i*.

$$POD_i = \frac{A_{ii}}{R_i}, where\ i = 1,\ \ldots\ , k$$

Sometimes it is useful to combine two or more categories from a contingency table into a single category and compute a POD for the new category. For a description of this type of specialized contingency table and the POD formula, see sections 3 and 3.1.

2.4 <u>False Alarm Ratio (FAR)</u>. A *FAR* may be calculated for each individual category, *i,* of Table A-1. It measures the fraction of forecasts of category *i* that were incorrect. It gets its name "false alarm" from the times when category *i* is a rare or extreme event that may require a warning, watch or advisory.

$$FAR_i = \frac{C_i - A_{ii}}{C_i},\ where\ i = 1,\ \ldots\ , k$$

Sometimes it is useful to combine two or more categories from a contingency table into a single category and compute an FAR for the new category. For a description of this type of specialized contingency table and the FAR formula, see sections 3 and 3.2.

2.5 <u>Critical Success Index (CSI)</u>. A *CSI* may be calculated for each individual category, *i,* of Table A-1. It measures the forecaster's success in covering each event of category *i* with a correct forecast, A_{ii}, while also penalizing for incorrect forecasts of category *i*. It differs from the *POD* in that the *POD* doesn't penalize for incorrect forecasts.

$$CSI_i = \frac{A_{ii}}{R_i + C_i - A_{ii}},\ where\ i = 1,\ \ldots\ , k$$

Sometimes it is useful to combine two or more categories from a contingency table into a single category and compute a CSI for the new category. For a description of this type of specialized contingency table and the CSI formula, see sections 3 and 3.3.

2.6 <u>Generalized Skill Score (SS)</u>. This generalized skill score measures the fraction of possible improvement of the forecasts over some standard or test set of forecasts.

$$SS = \frac{NC - E}{N - E}, where:$$

$$NC\ (number\ correct) = \sum_{i=1}^{k} A_{ii}$$

and *E* represents some standard or test set of forecasts.

2.7 Heidke Skill Score (HSS). Sometimes the standard or test forecasts *(E)* from the generalized skill score (see section 2.6) are the values expected by chance and are computed from the marginal totals of the contingency table. One such score is the *HSS.*

$$HSS = \frac{NC - E}{N - E}, where:$$

$$NC\ (number\ correct) = \sum_{i=1}^{k} A_{ii}\ ; \quad E = \sum_{i=1}^{k} \frac{C_i R_i}{N}$$

A perfect Heidke skill score is one. Zero is indicative of no skill, and a negative score indicates skill worse than random forecasts. With three or more categories in the contingency table, Heidke only allows credit for categorical forecast hits along the diagonal of the contingency table, and therefore, does not penalize large categorical errors more than small categorical errors. This property rules out the possibility for granting "partial credit" to small forecast errors or "near hits." Also, correct forecasts of low frequency events are treated the same as correct forecasts of common events so the forecaster is not encouraged to forecast climatologically improbable (rare) events.

The CPC uses a version of the Heidke skill score for its main verification statistic. This is calculated by the formula:

$$HSS = \frac{NC - CH}{NT - CH} x100,$$

where NC is the total number of locations for which the forecast was correct, NT is the total number of locations for which a forecast was made, and CH is the number of locations which would be forecast correctly, on average, by chance. In a three class system (which is how all the CPC forecasts are characterized), one third of the locations are expected to be correct by chance. Thus if 99 locations are forecast, 33 are expected to be correctly forecast. This statistic results in scores of 100 if all locations are forecast correctly, zero if 33 are forecast correctly, and -50 if all locations are forecast incorrectly.

2.8 Peirce Skill Score *(PSS)*. The Pierce skill score (Peirce 1884), also known as the Hanssen–Kuipers discriminant (Hanssen and Kuipers, 1965) and the true skill statistic (Flueck 1987), is similar to the Heidke skill score. Peirce and Heidke differ only in how they estimate the number of correct forecasts that would be expected by chance in their respective denominators—the numerators of the two scores are identical. Both scores are equitable, which means that a perfect forecast (all correct) results in a score equal to one, and a no skill (random) forecast results in a score equal to zero. Negative scores are possible. With three or more categories in the contingency table, Peirce only allows credit for categorical forecast hits along the diagonal of the contingency table, and therefore, does not penalize large categorical errors more than small ones. This property rules out the possibility for granting "partial credit" to small forecast errors or "near hits." Also, with three or more categories in the contingency table, correct forecasts of low frequency events are treated the same as correct forecasts of common events so the forecaster is not encouraged to forecast climatologically improbable events.

$$PSS = \frac{NC - E}{N - E^*}, where:$$

$$NC\ (number\ correct) = \sum_{i=1}^{k} A_{ii} \qquad E = \sum_{i=1}^{k} \frac{C_i R_i}{N} \qquad E^* = \sum_{i=1}^{k} \frac{R_i R_i}{N}$$

2.9 Equitable Skill Scores (ESS).

2.9.1 Subjective Explanation. Skill scores are often used to evaluate multi-category forecasts with a single score. Equitability is a desirable property for a skill score because equitability has the following characteristics:

a. A set of perfect forecasts (all categorical hits) produces a score equal to one.

b. A set of randomly generated forecasts or a set of forecasts that always predicts the same forecast category results in a "no skill" score equal to zero.

While equitable skill scores, such as Heidke (section 2.7) and Peirce (section 2.8), are convenient (they can often be computed by hand), they only grant credit for categorical forecast hits. Therefore, with three or more categories in the contingency table, Peirce and Heidke do not penalize large categorical errors more than small ones, and this rules out the possibility of receiving partial credit for "near hits." Also, correct forecasts of low frequency events are treated the same as correct forecasts of very common events so the forecaster is not encouraged to forecast climatologically improbable (rare) events.

Gandin and Murphy (1992) developed a mathematical framework for computing equitable scores that allow for a system of graduated, partial credit that considers the size of each miss and the observed frequency of each category. While Gandin and Murphy allowed for forecast systems with a higher number of forecast categories, examples of systems with greater than three categories were beyond the scope of their work. Gerrity (1992) built upon Gandin and Murphy and derived a general set of formulas that place no upper limit on the number of categories allowed in the system. Gerrity's formulas must be applied to scoring forecasts of ordinal variables (order matters) with maximum and minimum values, e.g., temperature, wind speed, ceiling, and visibility. While high speed computation is necessary for the Gerrity formulas, they are relatively simple to program. The Gerrity ESS has been implemented operationally in the NWS and has the following reward/penalty characteristics:

a. A relatively small reward is given for correctly forecasting common events.

b. A large reward is given for correctly forecasting rare events.

c. A graduated reward/penalty system is used, whereby a large forecast error for a given category is penalized more than a small forecast error for that category.

d. Less penalty is assigned to an incorrect forecast of a rare event than a similar size error of a common event. "Near hits" of rare events often receive a modest reward.

The otherwise favorable property of giving large rewards for correct forecasts of rare events may make the score volatile, especially with very small sample sizes of the rare events. For example, if a particular event occurs on a very rare basis, the ESS may increase substantially due to just one additional correct forecast of that rare event. Therefore, the ESS is not the ideal score for data requests that include relatively small geographic areas and/or short periods of time with little variability in the element. It is also important to exercise care in defining categories in the first place to keep very rare events and volatile scores from becoming a foregone conclusion.

Depending upon the element being verified, the rarest categories tend to be either the lowest or highest categories of the contingency table. For example with wind speed and significant wave height, the rarest events tend to be the highest categories. With ceiling and visibility, the rarest events tend to be the lowest categories. The ESS Low/High Category Delta is defined as the increase that occurs in the ESS due to one additional forecast hit in the lowest/highest category whose event count is at least one. Whenever the ESS is used, the delta values should always be checked for potential score volatility. A delta value that is unacceptably high should lead the user of *Stats on Demand* to resubmit a data request for a larger geographic area and/or longer time frame. See the last two paragraphs of section 2.9.2 for the mathematical definitions of the delta values.

2.9.2 Mathematical Background. The probability matrix, **P**, comes from the **A** matrix (Table A-1), where all

$$p_{ij} = \frac{A_{ij}}{N} \quad ; \quad (i = 1, \ldots, k \text{ and } j = 1, \ldots, k)$$

The row totals of the **P** matrix comprise **p**, the climatological probability vector, ($p_1, p_2, \ldots, p_k$). The column totals of the **P** matrix comprise **q**, the forecast probability vector, ($q_1, q_2, \ldots, q_k$).

Gandin and Murphy (1992) describe what is meant by an "equitable skill score" for the evaluation of categorical forecasts. The general formula is

$$ESS = \sum_{i=1}^{k}\sum_{j=1}^{k} p_{ij}s_{ij}$$

Note that p_{ij} are the elements in the aforementioned **P** matrix, and s_{ij} are the elements of the reward-penalty matrix, also called the scoring matrix (**S**). When an appropriate climatology is used to populate the **S** matrix, a random set of forecasts yields an ESS equal to zero, and a perfect set of forecasts (i.e., only the diagonal of the **P** matrix is populated) yields an ESS equal to one.

Gerrity (1992) derived the following formulas for populating the **S** matrix in a *k*-category system.
These formulas are only appropriate for ordinal variables (i.e., the order of the categories matters) that are not periodic. Wind speed and ceiling height are examples of ordinal, non-periodic variables. Wind direction is an example of an ordinal, periodic variable for which the Gerrity solution is not appropriate because as an eight-category variable, wind direction cannot "miss" by more than four categories (a non-periodic variable expressed in terms of eight categories can miss by up to seven categories).

Gerrity defines *p(r)* as the relative frequency with which category *r* of an event is observed in a large sample of forecasts and then defines *D(n)* and *R(n)*:

$$D(n) \equiv \frac{1-\sum_{r=1}^{n} p(r)}{\sum_{r=1}^{n} p(r)} \qquad\qquad R(n) = \frac{1}{D(n)}$$

D(n) is the ratio of the probability that an observation falls into a category with an index greater than *n* to the probability that it falls into a category with an index less than or equal to *n*; *R(n)* is the reciprocal of this ratio of probabilities. In terms of *D* and *R*, Gerrity expresses the elements of a *k*-category equitable **S** matrix in the following manner:

$$s_{m,n} = \frac{1}{k-1}\left[\sum_{r=1}^{m-1} R(r) + \sum_{r=m}^{n-1}(-1) + \sum_{r=n}^{k-1} D(r)\right] \quad ; \quad n = (1,\dots,k)$$

$$s_{n,n} = \frac{1}{k-1}\left[\sum_{r=1}^{n-1} R(r) + \sum_{r=n}^{k-1} D(r)\right] \;;\; 1 \le m < k, \quad m < n \le k$$

$$s_{n,m} = s_{m,n} \;;\; 2 \le n \le k, \quad 1 \le m \le n$$

Burroughs (1993), appendix B, section n, applies these general equations for populating the **S** matrix to specific *k*-category marine elements.

The **S** matrix is computed directly from the sample of the *Stats on Demand* data request. This practice has one major shortcoming; requests for verification data from relatively small, restrictive samples will tend to produce volatile scores that fluctuate due to random changes in the data set. Ironically, this problem is aggravated in these situations by the otherwise favorable ESS property of giving more weight to rare events. The following two paragraphs describe the measure used to help identify these situations.

Depending upon the element being verified, the rarest categories tend to be either the lowest or highest categories of the contingency table. To help the user of *Stats on Demand* test the ESS for volatility, one or both of the following "deltas" are calculated and listed in the verification reports with the ESS:

$$\delta_{low} = \frac{s_{aa}}{N}$$

$$\delta_{high} = \frac{s_{bb}}{N}$$

where δ_{low} is defined as the increase that occurs in the ESS due to one additional forecast hit in *a,* the lowest category in the contingency table whose total event count is at least one, and δ_{high} is defined as the increase that occurs in the ESS due to one additional forecast hit in *b,* the highest category in the contingency table whose total event count is at least one.

The user of *Stats on Demand* can easily calculate the delta for any intermediate category, *i*, in the contingency table by dividing the weight given in the reward-penalty matrix for a correct forecast in the *i*th category (s_{ii}) by the total sample size (N). The user of the ESS is strongly encouraged to pay close attention to the delta value provided with a particular score for an estimate of score volatility. If the score is too volatile for the user's tolerance, re-compute the score for a larger, less restrictive area in space and time.

3. <u>Specialized Contingency Table</u>. The following contingency table (Table A-2) may be used when only two outcomes (yes or no) exist for a given event or forecast, e.g., tornadoes. The number of correct forecasts for the specific event is given by *A*. The number of events observed but not forecast is given by *B*. The number of forecasts which did not verify is represented by *C*. The number of times the specific event was neither forecast nor observed is represented by *X.*

Table A-2 may be obtained from Table A-1 by combining multiple categories of Table A-1. For example with marine forecasts, sustained wind speeds are divided into seven categories. Define sustained wind speeds equaling or exceeding 28 knots (categories 6 and 7) as the "yes" outcome for a strong wind forecast or event. In this case, the "no" outcome is all sustained wind speeds less than 28 knots (categories 1 through 5 combined). The result is two categories (yes and no).

Table A-2. Specialized Contingency Table

		Forecasts	
		Yes	No
Events	Yes	*A*	*B*
	No	*C*	*X*

The scores most frequently computed from this table are:

3.1 Probability of Detection *(POD)* is the fraction of actual events *(A+B)* correctly forecast *(A)*. In the case of warnings, the *POD* is computed from the event database and is the number of warned events divided by the total number of events. The more often an event is correctly forecast, the better the score. The best possible score is 1, the worst possible score is 0.

$$POD = \frac{A}{A + B}$$

If *(A+B)* is the total number of events, e.g. turbulence or icing, sometimes it is useful to compute the *POD* of null events, e.g., no turbulence or no icing. Thus the *POD* of null events *(POD[N])* is the probability of null events that were forecast correctly. An alternative name for this statistic is the probability of null events *(PON).* The formula is

$$POD[N] = PON = \frac{X}{X + C}$$

3.2 False Alarm Ratio *(FAR)* is the fraction of all forecasts *(A+C)* which were incorrect *(C)*. In the case of warnings, the *FAR* is computed from the event database and is the number of false alarms (unverified warnings) divided by the total number of warnings. The more often an event is forecast and does not occur, the worse the score. The best possible score is 0, the worst possible score is 1.

$$FAR = \frac{C}{A + C}$$

The *POD* and *FAR* are most often used in the verification of watches and warnings. However, it is possible to apply the *POD* and *FAR* to many events and forecasts related to public and aviation elements. Two examples are the *POD* for ceilings below 1000 feet and the *FAR* for forecasts of freezing rain.

Over-forecasting an event will achieve a high *POD* but at the expense of a high *FAR*. Overall success can be expressed by the critical success index *(CSI)*.

3.3 Critical Success Index is the ratio of correct forecasts *(A)* to the number of events *(A+B)* plus the number of incorrect forecasts *(C)*.

$$CSI = \frac{A}{A + B + C}$$

The best possible score is 1, the worst is 0. The relationship among *POD*, *FAR*, and *CSI* can be expressed as follows:

$$CSI = [(POD)^{-1} + (1 - FAR)^{-1} - 1]^{-1}$$

In the case of severe thunderstorm watches and warnings, the value of *A* varies depending upon whether it is taken from the warning or the event database. This is true because multiple events within a single county are sometimes counted as separate events in the event database, whereas only one warning can be in effect for a particular county at the same time. For this reason, the number of warned events in the event database, denoted below as A_e, may exceed the number of verified warnings in the warning database, denoted below as A_w. Using these conventions, the definitions of *POD* and *FAR* are

$$POD = \frac{A_e}{A_e + B}$$

$$FAR = \frac{C}{A_w + C}$$

Given these expressions for *POD* and *FAR* and the *CSI* formula, expressed in terms of *POD* and *FAR*, the *CSI* becomes:

$$CSI = \frac{A_w A_e}{A_w A_e + A_w B + A_e C}$$

4. Scores Computed for Specific Forecast Elements. Other scores may be computed, where N = number of cases; f_i = the *ith* forecast, and o_i = the *ith* observation (matching the forecast).

4.1 Temperature, Wind Speed and Direction, and Wave Height. Scores frequently computed for forecasts of temperature, wind speed and direction, and wave height include:

a. Mean Error *(ME)* indicates whether collective forecast values were too high or too low. This is also called the mean algebraic error.

$$ME = \frac{1}{N}\sum_{i=1}^{N}(f_i - o_i)$$

b. Mean Absolute Error *(MAE)* measures error without regard to the sign (whether positive or negative).

$$MAE = \frac{1}{N}\sum_{i=1}^{N}| f_i - o_i |$$

c. Root Mean Square Error *(RMSE)* weights large errors more than the MAE.

$$RMSE = \sqrt{\frac{1}{N}[\sum_{i=1}^{N}(fi - oi)^2]}$$

d. Measuring Errors Against Some Standard. The above measures of accuracy *(ME, MAE, RMSE)* may also be computed for some forecast standard, such as Model Output Statistics *(MOS)* guidance, climatology *(CLI)*, or persistence *(PER)*. For example, the *MAE* for *MOS* guidance forecasts (m_i) is

$$MAE_{MOS} = \frac{1}{N}\sum_{i=1}^{N}| m_i - o_i |$$

Forecast skill is determined by measuring the improvement of forecasts over a forecast standard. For example, the *MAE* may be used to compute the percent improvement of forecasts over *MOS*, $I(MAE)_{MOS}$.

$$I(MAE)_{MOS} = \frac{MAE_{MOS} - MAE}{MAE_{MOS}} x100$$

Other examples include $I(RMSE)_{MOS}$, $I(MAE)_{CLI}$, and $I(RMSE)_{PER}$.

4.2 Probability of Precipitation. Scores typically computed for probability of precipitation verification include:

a. Brier Score *(BS)* measures the mean square error of all PoP intervals forecast. The standard NWS Brier score, defined below, is one-half the original score defined by Brier (1950).

$$BS = \frac{1}{N}\sum_{i=1}^{N}(f_i - o_i)^2$$

where, f_i = forecast probability for the *ith* case, o_i = observed precipitation occurrence (0 or 1), and N = the number of cases.

b. Climatological Brier Score *(BS_{CLI})* is an application of the Brier score to forecasts, c_i, consisting of climatic relative frequencies, *RF* (see below).

$$BS_{CLI} = \frac{1}{N}\sum_{i=1}^{N}(c_i - o_i)^2$$

c. Improvement over Climate Based on Brier Score *($I(BS)_{CLI}$)* measures the improvement gained from actual forecasts versus climatological values.

$$I(BS)_{CLI} = \frac{BS_{CLI} - BS}{BS_{CLI}} x100$$

d. MOS Brier Score *(BS_{MOS})* is analogous to BS_{CLI}, except the Brier score is computed for MOS forecasts.

$$BS_{MOS} = \frac{1}{N}\sum_{i=1}^{N}(m_i - o_i)^2$$

where, m_i = MOS guidance probability for the *ith* case. MOS guidance probabilities *(m_i)* are forecast to the nearest 0.01; however for NWS PoP verification, the m_i values are rounded to one of the following values: 0, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, and 1.0.

e. Improvement over MOS Based on Brier Score *($I(BS)_{MOS}$)* is analogous to $I(BS)_{CLI}$, except this score measures the improvement of the forecast over MOS.

$$I(BS)_{MOS} = \frac{BS_{MOS} - BS}{BS_{MOS}} x100$$

f. Relative Frequency of an Event *(RF)* is the fraction of the time an event occurred.

$$RF = \frac{1}{N}\sum_{i=1}^{N} o_i$$

g. Reliability, a measure of bias, compares the number of forecasts of an event with the observed relative frequency of the event. The reliability may be determined overall or by forecast interval, e.g., 10 percent PoP intervals or (0, 5, 10, 20, 30, . . ., 80, 90, 100.

$$\frac{1}{N}\sum_{i=1}^{N} f_i \quad \text{compared with} \quad \frac{1}{N}\sum_{i=1}^{N} o_i \ ,$$

where, N is the total number of events or the number of events in the interval. If the number of forecasts of the event or interval is larger (smaller) than the observed relative frequency of the event or interval, the event or interval was overforecast (underforecast).

4.3. QPF.

a. Bias, Threat Score, POD, and FAR, when applied to QPF verification, are computed from gridded data for specific precipitation amount thresholds, e.g. 0.01 inch, 0.25 inch, 0.50 inch, 1.00 inch, etc. Bias *(B)* and Threat Score *(TS)* (Gilbert 1884; Junker et al. 1989; Schaefer 1990) (also known as the *CSI*) are defined as follows:

$$B = \frac{F}{O}$$

$$TS = CSI = \frac{H}{F + O - H}$$

where, F is the number of points forecast to have at least a certain amount (threshold) of precipitation, O is the number of points observed to have at least the threshold amount, and H is the number of points with correct forecasts for that threshold of precipitation. When the bias is less [greater] than unity for a given threshold, the forecast is under [over] forecasting the areal coverage for that amount.

Geometrically, the threat score for a given threshold amount represents the ratio of the correctly predicted area to the threat area. Threat area is defined as the envelope of forecast and observed areas for that threshold. A perfect forecast yields a threat score of one, and a forecast with no areas correctly predicted receives a zero. The threat score, therefore, provides a measure of how accurately the location of precipitation is forecast within the valid period of the forecast. To receive a high threat score, forecast precipitation must be accurate—both spatially and temporally. For example, if a 1.00-inch isohyet is forecast, and all the observed rainfall within that area ranges from 0.8 to 0.99 inch, the forecaster's 1.00-inch threat score would be zero. However, the 0.8 to 0.99 inch area would favorably affect the 0.5-inch threat score. Also, a forecast area that is adjacent to an observed area with no overlap produces a zero threat score, and forecasts that are incorrect by just a couple of hours may receive little or no credit. Closely related to the threat score are *POD* and *FAR* which are expressed as:

$$POD = \frac{H}{O}$$

$$FAR = \frac{F - H}{F}$$

b. Equitable threat score *(ETS)* (Messinger 1996) is similar to the threat score except the expected number of hits in a random forecast, *E,* is subtracted from the numerator and denominator:

$$ETS = \frac{H - E}{F + O - H - E}$$

where $E=FO/N$, and N is the number of points verified. E is substantial for low precipitation categories, i.e., 0.10 inch or less in 24 hours, small at intermediate categories, and negligible for high categories, i.e., 1 inch or more in 24 hours.

4.4 Ceiling Height and Visibility. The Log Score *(LS)* is used for verifying ceiling height and visibility forecasts. It emphasizes accuracy in the more critical lower ceiling height and visibility ranges.

$$LS = \frac{50}{N}\sum_{i=1}^{N} \left| Log_{10}\left(\frac{f_i}{o_i}\right) \right|$$

Where f_i is the category of the *ith* forecast and o_i is the category of the *ith* observation. Note, f_i and o_i may also be used to represent the actual respective forecast and observed values of the element (i.e., ceiling height in feet, visibility in statute miles). Persistence is often used as the reference standard for evaluating ceiling height and visibility forecasts. The last hourly observation available to the forecaster before dissemination of the terminal aerodrome forecast defines the persistence forecasts of ceiling height and visibility to which the TAFs are compared.

4.5 Aviation Weather Center (AWC) Verification Statistics. The following statistics are used for verifying AWC forecasts:

a Probability of Detection *(POD)*. Same as section 3a of this appendix.

b. False Alarm Ratio *(FAR)*. Same as section 3b of this appendix.

c. Probability of Detection of "No" Observations *(POD[N])* is an estimate of the proportion of "no" observations that were correctly forecast (i.e., PIREPs which include reports such as negative icing or negative turbulence). An alternative name for this statistic is the probability of null events *(PON)*. Based on the contingency table presented in section 3 of this manual,

$$POD[N] = PON = \frac{X}{X + C}$$

d. Percent Area *(% Area)* is the percentage of the forecast domain's area where the forecast variable is expected to occur. It is the percent of the total area with a YES forecast.

e. Percent Volume *(% Vol)* is the percentage of the forecast domain's volume where the forecast variable is expected to occur. It is the percent of the total volume with a YES forecast.

5. References.

Brier, G.W., 1950: Verification of forecasts expressed in terms of probability. Monthly Weather Review, **78**, 1-3.

Burroughs, L.D., 1993: National marine verification program - verification statistics. OPC Technical Note/NMC Office Note No. 400, National Weather Service, NOAA, U.S. Dept. of Commerce, 48 pp.

Burroughs, L.D., 2002: Verification scores from performance matrices—a short tutorial. Personal communication, NOAA, National Weather Service, National Centers for Environmental Prediction, Environmental Modeling Center, Ocean Modeling Branch (W/NP21).

Flueck, J.A., 1987: A study of some measures of forecast verification. *Preprints, 10th Conference on Probability and Statistics in the Atmospheric Sciences.* Edmonton, AB, Canada, American Meteorological Society.

Gandin, L.S., and A.H. Murphy, 1992: Equitable skill scores for categorical forecasts. Monthly Weather Review, **120**, 361-370.

Gerrity, J.P., 1992: A note on Gandin and Murphy's equitable skill score. *Mon. Wea. Rev.,* **120,** 2709-2712.

Gilbert, G.F., 1884: Finley's tornado predictions. American Meteorological Journal, **1**, 166-172.

Hanssen, A.W. and W.J.A. Kuipers, 1965: On the relationship between the frequency of rain and various meteorological parameters. *Mededeelingen en Verhandelingen,* Royal Netherlands Meteorological Institute, **81.**

Hughes, L.A., 1980: Probability forecasting - reasons, procedures, problems. NOAA

Technical Memorandum NWS FCST 24, National Weather Service, NOAA, U.S. Department of Commerce, 84 pp.

Junker, N.W., J.E. Hoke, and R.H. Grumm, 1989: Performance of NMC's regional models. Weather and Forecasting, 4, 368-390.

Livezey, R.E., 2003: Categorical events (chapter 4). *Forecast Verification: A Practitioner's Guide in Atmospheric Science.* Edited by I.T. Jolliffe and D.B. Stephenson, John Wiley and Sons, Ltd., 240 pp.

Messinger, F., 1996: Improvements in precipitation forecasts with the Eta regional model at the National Centers for Environmental Prediction: The 48-km upgrade. Bulletin of the American Meteorological Society, 77, 2637-2649.

Peirce, C.S., 1884: The numerical measure of the success of predictions. *Science,* **4,** 453-454.

Schaefer, J.T., 1990: The critical success index as an indicator of warning skill. Weather and Forecasting, **5**, 570-575.

Wilks, D.S., 1995: Statistical Methods in the Atmospheric Sciences. Academic Press, San Diego, CA, 467 pp.

APPENDIX B – Glossary of Contractions and Terms

AOP	Annual Operating Plan
CFW	Coastal Flood Warning
C-MAN	Coastal-Marine Automated Network (hourly weather observations)
CONUS	Contiguous United States
CPC	Climate Prediction Center
CSI	Critical Success Index, see Appendix A, section 3.3.
CWF	Coastal Waters Forecast
GFS	Global Forecast System Model
EF Scale	Enhanced Fujita Scale
ESS	Equitable Skill Score
FAR	False Alarm Ratio, see Appendix A, section 3.2
FFW	Flash Flood Warning
FLW	Flood Warning
GLF	Great Lakes Open Lake Forecast
HPC	Hydrometeorological Prediction Center
HRAP	Hydrologic Rainfall Analysis Project (frequently used 4-km grid)
HSS	Heidke Skill Score
LAMP	Local AWIPS MOS Program
LST	Local Standard Time
MAE	Mean Absolute Error, see appendix A, section 4.1
METAR	Meteorological Aviation Reports
MDL	Meteorological Development Laboratory
ME	Mean Error (algebraic), see appendix A, section 4.1
MOS	Model Output Statistics
MVF	Marine Verification Forecast (coded)
NAM	North American Mesoscale Model
NDFD	National Digital Forecast Database
NFDRS	National Fire Danger Rating System
NFWOC	National Fire Weather Operations Coordinator
NHC	National Hurricane Center
NPMC	National Performance Management Committee
NPVU	National Precipitation Verification Unit
NSH	Near Shore Forecast (Great Lakes)
OCWWS	Office of Climate, Water and Weather Services
OFF	Offshore Forecast
OPC	Ocean Prediction Center
PFM	Point Forecast Matrix (coded public forecast at points)
POD	Probability of Detection, see appendix A, section 3.1
PoP	Probability of Precipitation
PROB	Probabilistic Forecast in a TAF
PSS	Peirce Skill Score
QPE	Quantitative Precipitation Estimate (past precipitation analysis)

QPF	Quantitative Precipitation Forecast
RFC	River Forecast Center
RFW	Red Flag Warning
RMSE	Root Mean Square Error, see appendix A, section 4.1
RTMA	Real Time Mesoscale Analysis, provided hourly for select elements
SMW	Special Marine Warning
SPECI	Special Aviation Weather Reports
SVR	Severe Thunderstorm Warning
TAF	Terminal Aerodrome Forecast
TCM	Tropical Cyclone Forecast/Advisory
TOR	Tornado Warning
TPC	Tropical Prediction Center
TEMPO	Temporary Forecast Conditions in a TAF
VCTS	Thunderstorms in the vicinity (within a 5- to 10-mile radius) of the aerodrome
VTEC	Valid Time and Event Code

Accuracy is a measure of how much a forecast agrees with the event or element being forecast. The smaller the difference between the forecast and observation, the greater the accuracy.

***Change* –** For Terminal Aerodrome Forecasts (TAF), this term is used in defining condition variability for the Operational Impact Forecast.

***Area Corrected Lead Time* –** For flash flooding, the warned area lead time is multiplied by the percentage of the area warned.

Hit – A correct forecast, as defined by a contingency table or some forecast error threshold value. In a binary situation (yes or no), the term "hit" is reserved for correctly forecasting the yes event. The yes event may be rare or

***Lead Time* –** The amount of advance notice provided by a watch or warning concerning some operationally significant or life-threatening weather phenomenon. Negative lead times (when the warning is issued after the event is first observed) are recorded as zero.

Operational Impact Forecast – A two-step process for determining whether to base the verification of a TAF on (a) the prevailing forecast or (b) the TEMPO or PROB forecast, whichever may be in effect at a given projection time. For more detail, see section 6.1.6.2.

***Percentage of the Area Warned* –** With flash flood warnings, the area of reported flash flooding and the forecast area of the warning are overlaid to compute the percentage of the event area that was warned.

Storm Data - NOAA's official publication which documents the occurrence of storms and other significant weather phenomena having sufficient intensity to cause loss of life, injuries, significant property damage, disruption to commerce, and other noteworthy meteorological events.

Skill is a measure of improvement of a forecast over an established standard. Examples of standards often used for comparison include the climatological frequency (or value), persistence, or forecasts made by another process (e.g., model output statistics). The greater the improvement, the greater the skill.

Timing Error **–** In warning verification, the timing error is defined as the event beginning time minus the forecast start time in the warning.

Warned Area Lead Time **–** For a flash flood event, the warned area lead time is calculated by subtracting the warning issuance time from the time when the event began. Negative lead times are set to zero.