Department of Commerce • National Oceanic & Atmospheric Administration • National Weather Service
NATIONAL WEATHER SERVICE INSTRUCTION 10-1601
NOVEMBER 3, 2005

Operations and Services

Performance, NWSPD 10-16

VERIFICATION PROCEDURES

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

OPR: OS52 (Charles Kluepfel) **Certified by:** OS5 (Kimberly Campbell)

Type of Issuance: Routine

SUMMARY OF REVISIONS: This directive supersedes National Weather Service Instruction 10-1601, dated April 10, 2004. The following changes have been made to this directive:

1) Introductory information on the legacy verification systems for public (section 1.1.1) and terminal aerodrome (section 6.1.1) forecasts has replaced outdated information in these sections.

2) The verification of fire weather forecasts (section 1.4) and red flag warnings (section 1.5) has been added.

3) Monthly preliminary reporting requirements for tornado and flash flood warning verification statistics have been discontinued and respectively removed from sections 2.1 and 4.1.

4) Rule 2, implemented January 1, 2002, for short-fused warnings, has been discontinued for all tornado/severe thunderstorm (section 2.1.2) and special marine (section 3.3.2) warnings issued after February 28, 2005. Once VTEC is implemented for flash flood warnings (section 4.1.2), Rule 2 will also be discontinued for all flash flood warnings issued after the change.

5) The time of warning issuance for tornado, severe thunderstorm and special marine warnings is taken from the Valid Time and Event Code (VTEC) line (sections 2.1.3 and 3.3.3).

6) The National Digital Forecast Database (NDFD) quantitative precipitation forecasts (QPF) are now verified. All QPF verification is now done at 4- and 32-km resolution (section 5).

7) *Stats on Demand* (section 6.2) has replaced *Aviation Verify* (section 6.3) as the official terminal aerodrome forecast (TAF) verification tool.

8) Tropical cyclone verification procedures (section 7) have been updated.

________signed______________________________10/17/05___________
Dennis McCarthy Date
Director, Office of Climate, Water, and Weather

Verification Procedures

Table of Contents: Page

Appendices

1. Public and Fire Weather Forecast Verification Procedures.

1.1 Legacy Public Forecast Verification.

1.1.1 Introduction. The National Oceanic and Atmospheric Administration's (NOAA) National Weather Service (NWS) maintains a long-term record of public forecast verification statistics for the following elements: maximum and minimum temperature and 12-hour probability of precipitation (PoP). Quality controlled data are archived in a centralized database and are available to NWS employees through query, using the *Stats on Demand* feature of the NWS Verification Web Page. A list of verification sites, known as the legacy network, appears on this Web site, and data are available for most of these locations from April 1966 to September 2004. Each Weather Forecast Office (WFO) has at least one legacy verification site in its area of forecast responsibility, and a list of these sites appears on the NWS Verification Web Page. The legacy system no longer imports new data. For information about recent public forecast verification (2004 and later) and the modernized verification systems, see sections 1.2 and 1.3.

The NWS Verification 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 and generates verification statistics customized to the user's request. The user requests data for one public weather element and a single guidance product for one or more:

a. months;

b. model cycles;

c. projections; and

d. verification sites (single site, multiple site, regional, or national).

The Meteorological Development Laboratory (MDL) Evaluation Branch also operates and maintains an interactive Web page with public forecast verification statistics customized to the user's request and displayed in the form of scatter plots and reliability diagrams.

1.1.2 <u>Elements</u>. Projections for public elements verified at specific points in time (e.g., precipitation type) are defined as the number of hours elapsed since the appropriate model cycle initialization used for the forecast, 0000 Universal Coordinated Time (UTC) for the early morning forecast package and 1200 UTC for the late afternoon forecast package.

a. <u>Max/Min Temperatures</u>.

(1) <u>Projections</u>: The first four 12-hour forecast periods are verified: 24-, 36-, 48-, and 60-hour forecasts.

(2) <u>WFO Forecasts and Guidance</u>: Daytime maximum (max) and nighttime minimum (min) temperatures were recorded in whole degrees Fahrenheit for the first four 12- or 13-hour forecast periods. Daytime is defined as 7 a.m. to 7 p.m. Local Standard Time (LST). Nighttime is defined as 7 p.m. to 8 a.m. LST.

(3) <u>Observations</u>: Daytime max and nighttime min temperatures, in degrees Fahrenheit, were inferred from the Meteorological Aviation Reports (METAR) with an algorithm (Beasley 1995) that used the 6-hour max/min temperatures (1xxxx and 2xxxx groups) and hourly temperature readings.

b. <u>Probability of Precipitation (PoP)</u>. Probability of 0.01 inch or greater liquid equivalent precipitation within a 12-hour period: 0000 to 1200 UTC and 1200 to 0000 UTC in the Contiguous United States and Alaska; 0600 to 1800 UTC and 1800 to 0600 UTC in Hawaii.

(1) <u>Projections</u>: The first three 12-hour forecast periods are included: 24-, 36-

and 48-hour forecasts.

(2) WFO Forecasts and Guidance: The following forecast percentages were recorded in the digital forecast matrices and verification database: {0, 5, 10, 20, 30, ..., 80, 90, 100}. Guidance PoPs, given to the nearest percent, were rounded to the nearest allowable value.

(3) Observations: From METAR, 12-hour precipitation amounts to the nearest hundredth of an inch were recorded.

1.2 Modernized Public Forecast Verification at Points. NWS public forecasts are verified from the point forecast matrices (PFM). Forecast elements are verified out to Day 7.

1.2.1 Verification Sites. Public forecasts are verified at all sites forecast in the point forecast matrices that issue routine METARs and Special Aviation Weather Reports (SPECI) unless a local WFO determines that a particular site is unrepresentative of its surroundings or inappropriate for verification. Cooperative (COOP) observer reports will be incorporated when the COOP program is modernized and produces data compatible with NWS forecasts and guidance. A list of all active verification sites is maintained on the NWS Verification Web Page. 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).

1.2.2 Data Input. Public forecast data come from the scheduled PFMs issued by each WFO twice a day at 0400 and 1600 Local Time. The latest 0400 (1600) LT PFM issued between 0000 and 0559 (1200 and 1759) LT, including corrections, are accepted. Amendments are not verified at this time. The guidance forecasts come from the alphanumeric Model Output Statistics (MOS) guidance products derived from the following models: North American Mesoscale (NAM) (formerly called the Eta), Global Forecast System (GFS), and Nested Grid Model (NGM). The latest version of guidance available at forecast issuance time is used. 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 Verification Web Page. Public forecast verification statistics are available back to January 2004. Requests for data prior to January 2004 must go to legacy public forecast verification (section 1.1).

1.2.3 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. For most elements, forecasts are made out to Day 7, totaling 13 or 14 projections in a single PFM. This is a departure from the legacy public forecast verification system which expresses projections in hours, uses the UTC clock, and measures the forecast projection from the model cycle used in the preparation of the forecast, i.e., 0000 UTC for the 4 a.m. LT forecast and 1200 UTC for the 4 p.m. LT forecast within most of the NWS.

1.2.4 Public Forecast Verification Reports. NWS employees access verification statistics from the *Stats on Demand* feature of the NWS Verification Web Page. This Web page is operated and maintained by the OCWWS Performance Branch. *Stats on Demand* accesses an interactive database 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; and

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

1.2.5 Elements.

a. Max/Min Temperatures. The forecast period for all daytime maximum (max) temperatures is defined as 7 a.m. to 7 p.m. LST. The forecast period for all nighttime minimum (min) 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 max and nighttime min temperatures are forecast in whole degrees Fahrenheit out to Day 7.

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

b. Probability of Precipitation (PoP). Probability of 0.01 inch or greater liquid equivalent precipitation within the following 12-hour periods: 0600 to 1800 Local Time (LT) and 1800 to 0600 LT. Note: The acronym LT refers to Local *Standard* Time when a given location observes Standard Time and to Local *Daylight* Time when that location observes Daylight Savings Time.

(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 guidance PoPs, forecast to the nearest percent, are rounded to the nearest allowable value. For comparison purposes, PoPs from the PFMs are matched to MOS guidance PoPs. Note, PFM PoPs are valid for the following periods: 0600 to 1800 LT and 1800 to 0600 LT. MOS PoPs are valid for 0000 to 1200 UTC and 1200 to 0000 UTC, except in Pacific Region where they are valid 0600 to 1800 UTC and 1800 to 0600 UTC. Hence, when PFM PoPs are matched to MOS PoPs, a 1- to 4-hour mismatch occurs, depending upon the local time zone. The forecast for the closest matching MOS forecast period is used, and no "correction factors" are applied to the MOS PoPs.

(3) Observations: From METARs, 12-hour precipitation amounts to the nearest hundredth of an inch are inferred. 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.

1.3 National Digital Forecast Database (NDFD) Verification. MDL verifies the NDFD out to Day 7. The following methods are used:

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

b. Grid-to-Grid. All grid points are verified from a 20 kilometer Rapid Update Cycle (RUC) analysis of available data interpolated to the NDFD 5-kilometer grid. These data are experimental.

The following elements are verified out to 7 days:

a. 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.

b. 12-hour PoP. Forecast periods are defined 0000-1200 and 1200-0000 UTC.

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

d. Dew point. Every 3 hours out to 72 hours; then every 6 hours out to 7 days.

e. 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 Web site operated by MDL.

1.4 Fire Weather Forecast Verification. This automated, national program is expected to be implemented around the beginning of calendar year 2006 for the Contiguous United States (CONUS). Forecasts and observations come from the fire weather product with the AWIPS product identifier (PIL) NMCFWOxx or NMCFWOxxx, where xx refers to one of the six NWS regions, and xxx refers to a specific WFO. All forecasts and observations are valid at 1300 LST but are issued as separate bulletins with the same product name. Observations are issued shortly after the daily observation time. Forecasts are issued approximately 22 hours prior to forecast valid time.

1.4.1 Verification Sites. A database of all possible verification sites is posted to the NWS Verification Web page.

1.4.2 Fire Weather Forecast Verification Reports. NWS employees access verification statistics from the *Stats on Demand* feature of the NWS Verification Web page. Data are only available for the CONUS. This Web page is operated and maintained by the OCWWS Performance Branch. *Stats on Demand* accesses an interactive database and generates verification statistics customized to the user's request. The user requests data with the following definitions and boundaries:

a. Element.

b. Temporal domain for statistical computations, i.e., start date and end date. 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, 1999, thru December 31, 2004.

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 will be 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. POD/FAR/CSI threshold window. For temperature, relative humidity, and wind speed, the user must specify the window of values from which the POD, FAR and CSI will be calculated. Examples: between 90ºF and 120ºF (temperature), between 30 and 100 mph (wind speed).

1.4.3 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.

1.5 Red Flag Warnings (RFW). Perform red flag warning verification manually at the WFO.

1.5.1 Matching Warnings and Events. Treat each fire weather zone as a separate verification area. Therefore, count a warning covering three zones as three warned areas or three warnings. For verification purposes, define an event as a situation when weather conditions meet or exceed the local warning criteria set by the NWS region or local Annual Operating Plan (AOP).

Record warnings and events in separate databases. All listings in the event database must meet warning criteria. Count one verified warning and one warned event whenever the event occurs in a warned zone. Count one unwarned event if the event occurs in a zone with no warning. Record one unverified warning for each warned zone that does not experience an event.
The majority of Red Flag Warnings are based on wind and humidity criteria, or some index based on these parameters. However, there are warnings that are issued due to the occurrence of dry lightning in some areas. These two events can have big differences in lead time and Probability of Detection (POD), False Alarm Rate (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 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. Until NWSI 10-404 is updated, procedures in paragraphs 1.5.1 and 1.5.5 of this directive override reporting procedures stated in NWSI 10-404.

1.5.2 Determining Events and Warnings. Events are determined by querying the weather observations for a zone to see if warning criteria were met. Events are not determined by querying users to determine if they feel you hit or missed warnings.

Warnings are issued based on two factors: weather and fuel conditions. The former is determined by the forecaster and the later is determined by the user. If warning criteria is met as determined by the weather, but a warning was not issued because the user determined that the fuels were not critical enough to warrant a warning, then this does not count as a missed event.

Warnings are verified based on meeting weather criteria. Warnings are not verified based on number of fire starts.

1.5.3 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.

1.5.4 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. Set negative values to zero. If a zone experiences an event meeting warning criteria with no warning 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.

1.5.5 Regional Reports. The 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 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:

Number of correct warnings: A
Number of warnings that did not verify: B
Number of unwarned events: C
POD: A/(A+C)
FAR: B/(A+B)
CSI: A/(A+B+C).

Until NWSI 10-404 is updated, procedures in paragraphs 1.5.1 and 1.5.5 of this directive override reporting procedures stated in NWSI 10-404.

1.6 Winter Storm Warnings. Perform winter storm and high wind warning verification manually at the WFO. For verification purposes, treat all winter storm, heavy snow, blizzard, heavy sleet, and ice storm warnings and events generically as **winter storm** warnings and events. See Table 1 for a listing of the events, as listed in *Storm Data,* that may be used to verify a winter storm warning.

1.6.1 Matching Warnings and Events. Treat each public forecast zone as a separate verification area. Therefore, count a warning covering three zones as three warned areas or three warnings. For verification purposes, define an event as a situation when weather conditions meet or exceed the local warning criteria set by the NWS region (e.g., 4 inches or more of snow in 12 hours or less). This is consistent with the guidelines for preparing *Storm Data.*

Table 1. *Storm Data* entries that verify winter weather and high wind warnings.

Warning Type	***Storm Data* entries that verify the warning**	***Storm Data* entries that do not verify the warning**
Winter storm, Heavy snow, Blizzard, Sleet, or Ice Storm	Winter storm, Heavy snow, Blizzard, Sleet Storm, or Ice Storm	Winter Weather/Mix
High Wind	High Winds	Strong Winds

Record warnings and events in separate databases. All listings in the event database must meet warning criteria. Do not record multiple events for a single zone. Count one verified warning and one warned event whenever the event occurs in a warned zone.

Count one unwarned event if the event occurs in a zone with no warning. Record one unverified warning for each warned zone that does not experience an event.

1.6.2 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. Count each extension in time of a zone already warned as a new warning only if the issuance time of the extension preceded the time warning criteria were first met. Count no more than one verified warning per event. Table 2 illustrates a hypothetical time extension of a heavy snow warning for a given zone. If the snow starts falling at 1400L Monday, but warning criteria are not met until 0000L Tuesday (marked as "X" on the diagram), Warning 1 counts as an unverified warning, and Warning 2 counts as a verified warning because it was issued prior to 0000L Tuesday. Also, record one warned event. Table 3 gives some examples of area and temporal extensions.

Table 2. Time extensions of warnings.

```
0500L                         1700L
Monday                        Monday
|-----------warning 1-----------|

                              1600L                                   1200L
                              Monday                                  Tuesday
                              |------------warning 2 (time extension)----------|

                                            0000L
                                            Tuesday
                                               X
```

Table 3. Examples of extensions to winter weather warnings.

WFO actions	**Weather that occurred**	**Verification results**
A heavy snow warning is issued for zones A and C on Day 1 at 1100 LST, valid from Day 1/1800 LST until Day 2/0600 LST. On Day 1 at 1700 LST, the warning is changed to a snow advisory in zone C.	Warning criteria were not met in zones A or C during the valid period (Day 1/1800 LST to Day 2/0600 LST).	Score one unverified warning (false alarm) for zone A, and one unverified warning (false alarm) for zone C.
On Day 2 at 0500 LST, the heavy snow warning for zone A was extended in time until Day 2 at 1700 LST. At the same time, the warning for zone C was changed to an advisory.	Warning criteria were met in zones A and C on Day 2 at 1500 LST.	Score one verified warning and one warned event with a lead time equal to 10 hours for zone A. Score one unwarned event (zero lead time) for zone C since the previous warning had been changed to an advisory.
The advisory for zone C was not upgraded to a warning because the event was ending on Day 2 in the late afternoon. On Day 2 at 1600 LST, the warning for zone A, which had already verified one hour earlier, was extended until	Accumulating snow in zone C ceased on Day 2 at 1600 LST. Accumulating snow continued until Day 3/0200 LST in zone A. Zone B, which had received all rain during the morning of Day 2, experienced a rapid changeover to snow at 1200	The time extension issued at 1600 LST for zone A is not scored since it was issued *after* the time warning criteria were met. The area extension for zone B is treated as a new warning for zone B. Score one verified

Day 3 at 0600 due to accumulating snow continuing in zone A. At 1800 LST, the area coverage of the heavy snow warning for zone A, valid until Day 3/ 0600 LST, was extended to zone B.	LST. Warning criteria were met in zone B at 1500 LST (Day 2). Accumulating snow ceased in zone B on Day 3 at 0200 LST.	warning and one warned event with zero lead time.

1.6.3 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. Set negative values to zero. If a zone experiences an event meeting warning criteria with no warning 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.

1.6.4 Reports. The regional headquarters will report the fiscal year's verification statistics to the OCWWS Performance Branch using the format in Table 4. All reports for the current fiscal year are due by the fifth working day of March (October through December data), June (October through March data), September (October through June data), and December (entire previous fiscal year). The OCWWS Performance Branch subsequently collates regional data into national summaries.

Table 4. Example of format for reporting winter storm and high wind verification statistics.

Verification Statistics **Fiscal Year ______________** **Date of Report ____________** **Region ____________**	**Winter Storms**	**High Winds**
Number of Warnings Issued		
Number of Verified Warnings		
Number of Unverified Warnings		
Number of Events		
Number of Events with Warnings		
Number of Events without Warnings		
Average Lead Time		
Probability of Detection (POD)		
False Alarm Ration (FAR)		

Critical Success Index (CSI)		

1.7 High Wind Warnings. Perform high wind warning verification manually at the WFO.

1.7.1 Matching Warnings and Events. Treat each public forecast zone as a separate verification area. Therefore, count a warning covering three zones as three warned areas or three warnings. For verification purposes, define an event as a situation when weather conditions meet or exceed the local warning criteria set by the NWS region (e.g., winds of 58 mph or greater for any duration). This is consistent with the guidelines for preparing *Storm Data.* As indicated in Table 1, a report of "high winds" in *Storm Data* verifies a high wind warning, but a report of "strong winds" means reported wind speeds fell short of high wind warning criteria.

Record warnings and events in separate databases. All listings in the event database must meet warning criteria. Do not record multiple events for a single zone. Count one verified warning and one warned event whenever the event occurs in a warned zone. Count one unwarned event if the event occurs in a zone with no warning. Record one unverified warning for each warned zone that does not experience an event.

1.7.2 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. Count each extension in time of a zone already warned as a new warning if the issuance time of the extension preceded the time warning criteria were first met.

1.7.3 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. Set negative values to zero. If a zone experiences an event meeting warning criteria with no warning 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.

1.7.4 Reports. The regional headquarters report the fiscal year's verification statistics to the OCWWS Performance Branch using the format in Table 4. All reports for the current fiscal year are due by the fifth working day of March (October through December data), June (October through March data), September (October through June data), and December (entire previous fiscal year). The OCWWS Performance Branch subsequently collates regional data into national summaries.

2. Severe Weather Verification Procedures. This section describes the verification of all severe thunderstorm and tornado watches and warnings.

2.1 Warning Verification. The OCWWS Performance Branch is responsible for the operation and maintenance of the automated severe weather warning verification program.

2.1.1 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 county. Therefore, for verification purposes, each county included in a warning statement is counted as a separate 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 groups severe thunderstorms and tornadoes together. The latter two methods are event specific—they treat non-tornadic severe thunderstorms and tornadoes as separate types of events. See Table 6 for illustration.

a. All Severe Thunderstorm and Tornado Verification (Generic). All severe thunderstorm and tornado data are treated as generic severe local storms. This means any tornado or severe thunderstorm warning may be verified by either a tornado or severe thunderstorm event. Likewise, to count as a warned event, any tornado or severe thunderstorm may be covered by either a tornado or severe thunderstorm warning.

b. Tornado only (TOR). A confirmed tornado is required to verify a tornado warning (using the TOR product). Likewise, to count as a warned event, a tornado must be covered by a tornado warning.

c. Severe Thunderstorm only. A severe thunderstorm warning (using the SVR product) is only verified by a non-tornadic severe thunderstorm, and a non-tornadic severe thunderstorm is only counted as a warned event if it is covered by a severe thunderstorm warning. Therefore, a tornado event does not verify a severe thunderstorm warning, and a tornado warning does not cover a non-tornadic severe thunderstorm event.

Table 6. *Storm Data* entries (events) used to verify local severe storm warnings.

<table>
<tr><th>Warning Type</th><th>Event Specific Verification
(bullets b. and c.)

Each warning in the left column is only verified by the corresponding event from the same line of this column. Each event in this column must be covered by the corresponding warning from the same line of the left column.</th><th>All Severe Thunderstorm and Tornado (Generic) Verification
(bullet a.)

Each warning in the left column is verified by any of the events in this column. Any event in this column must be covered by one of the warnings in the left column.</th></tr>
<tr><td>Severe thunderstorm</td><td>Non-tornadic severe thunderstorm</td><td rowspan="2">Non-tornado severe thunderstorm or tornado</td></tr>
<tr><td>Tornado</td><td>Tornado</td></tr>
</table>

All event data are automatically taken from the final *Storm Data* reports. Each severe weather warning may only be verified by a confirmed event meeting NWS warning criteria and occurring within the valid period and county represented by the warning. 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; therefore, only the first entry is recorded into the event database. 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*.

Warnings and events are recorded in separate databases. Whenever an event occurs in a warned county, the following are recorded: one verified warning and one warned event. One unwarned event is recorded for each event that occurs in a county with no warning. One unverified warning is counted for each warned county that does not experience an event.

2.1.2 Quality Assurance Rules. In an attempt to reduce the impact of erroneous short-fused warnings on users and, at the same time, more accurately measure the quality of NWS warnings, the OCWWS Performance Branch has developed a set of rules stating how these short-fused warnings are archived.

a. Rule 1 – How Warnings are Entered into the Database. All data imported into the warning database are taken directly from the warning. No data are entered into the database from any information other than that represented by the bold-faced parts of the warning sample in Table 7. Based on this rule, products issued with the improper coding may or may not be imported into the database. Several examples appear on the NWS Verification Web Page.

Table 7. Severe thunderstorm warning sample. The warning issuance and expiration times are taken from the VTEC line. The issuing WFO and warning type are checked for consistency with the top two lines of the warning. Inconsistent warnings are not counted for verification.

```
WUUS51 KWSH 010000
SVRWSH
DCC001-003-005-010100-
/O.NEW.KWSH.SV.W.0010.050101T0000Z-050101T0100Z/

BULLETIN - EAS ACTIVATION REQUESTED
SEVERE THUNDERSTORM WARNING
NATIONAL WEATHER SERVICE SILVER SPRING MD
700 PM EST WED JAN 1 2005

THE NATIONAL WEATHER SERVICE IN SILVER SPRING HAS ISSUED A

* SEVERE THUNDERSTORM WARNING FOR...
  WASHINGTON COUNTY IN THE DISTRICT OF COLUMBIA
  JEFFERSON COUNTY IN THE DISTRICT OF COLUMBIA
  REAGAN COUNTY IN THE DISTRICT OF COLUMBIA

* UNTIL 800 PM EST

* AT 700 PM CST...SILVER SPRING DOPPLER RADAR INDICATED A SEVERE
  THUNDERSTORM 2 MILES WEST OF ADAMS MORGAN...MOVING EAST AT 15 MPH.

THE SEVERE THUNDERSTORM IS CAPABLE OF PRODUCING...
  HAIL THE SIZE OF NICKELS

LAT...LON 3778 9752 3748 9752 3749 9724 3785 9724

$$
```

b. Rule 2 – Quality Assurance of Overlapping Warnings. This is valid for all warnings issued from January 1, 2002 to February 28, 2005. When two warnings for a given county overlap in time, the portion of the earlier warning that overlaps the second warning is removed. The expiration time of the first warning is changed to one minute before the issuance time of the second warning. Several examples appear on the NWS Verification Web Page.

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 time of warning issuance is subtracted from the time when an event meeting warning criteria was first reported in the county. The time of warning issuance is taken from Valid Time and Event Code (VTEC) line. Negative values are converted to 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. If one or more events occur in a county not covered by a warning, 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 tornado events with lead time greater than zero is also computed.

2.1.4 Display of Verification Statistics. NWS employees access verification statistics through the *Stats on Demand* feature of the NWS Verification Web Page. *Stats on Demand* accesses an interactive database that provides verification statistics customized to the user's request. The user may request data by:

a. type of warning;

b. one or more dates (select beginning and ending date);

c. one or more counties, WFOs, states, NWS regions, or the contiguous United States;

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

2.1.5 Backup Mode for Warnings. When a WFO goes into backup mode, warnings are still sorted by county, so all warnings issued by the backup office are attributed to the primary WFO.

2.2 Watch Verification. The Storm Prediction Center (SPC) is responsible for verifying the tornado and severe thunderstorm watches it issues. The area defined by a tornado or severe thunderstorm watch is defined as the verification area without regard to the number of counties affected. Weiss et al. (1980) describe 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 8 and 9.

Table 8. Coastal WFOs with marine responsibility.

Eastern Region WFOs	Western Region WFOs
Caribou, ME (CAR)	Seattle, WA (SEW)
Portland, ME (GYX)	Portland, OR (PQR)
Boston, MA (BOX)	Medford, OR (MFR)
New York City, NY (OKX)	Eureka, CA (EKA)
Philadelphia, PA (PHI)	San Francisco, CA (MTR)

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)	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 9. 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 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 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. A list of national marine verification sites appears on the NWS Verification Web Page.

3.1.3 Coded Forecast Format. Code the MVF in accordance with the format in Table 10. 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

MVF are intended only for the sensors of the buoys and C-MAN stations. See Table 11 for a sample CWF with the corresponding MVF.

Table 10. 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 11. 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
$$

A detailed explanation for each MVF entry is given below:

a. Forecaster Number (*nn*). Do NOT use comparative verification as an individual performance measure. However, once verification statistics become available for individuals, forecasters will review them for feedback, self-improvement, and knowledge.

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*). Enter the warning/advisory status with one of the 2-character abbreviations, explained below. This entry should represent the worst conditions expected during the appropriate valid period. For example, if a gale warning is issued due to forecast winds increasing to gale force during the second forecast period of the offshore forecast, enter "NO" advisory/warning for the 16- to 20-hour forecast and "GL" advisory/warning for the 28- to 32-hour forecast even though the headline near the beginning of the offshore forecast reads "Gale Warning."

A forecaster may enter a certain advisory/warning category in the MVF (e.g., gales), but forecast a mean speed (section 3.1.3.f, wind speed, *ff*) less than the minimum threshold for that warning category. Both entries are legitimate because the gale warnings are issued for the maximum forecast speed during the valid period, and the forecast wind speed is for the mean sustained wind speed expected during the valid period. The following 2-character entries are allowed:

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

(2) SC: Small craft advisories. These are the Small Craft Advisory, Small Craft Advisory for Hazardous Seas, Small Craft Advisory for Winds, and Small Craft Advisory for Rough Bar. Small craft advisories are only issued for the CWFs and NSHs.

(3) GL: Gale warning.

(4) ST: Storm warning.

(5) TS: Tropical storm warning.

(6) HR: Hurricane warning.

(7) HF: Warning for hurricane force winds in the absence of a hurricane.

e. Wind Direction (*dd*). Enter the 10 meter forecast mean wind direction for the MVF valid period in tens of degrees, e.g., enter "12" for a wind from 120 degrees. If a wind shift or variable winds are expected during the period, enter the forecast direction at the midpoint hour of the valid period (i.e., 0600 or 1800 UTC). If the wind direction is less than 100 degrees, place a zero in the tens digit, e.g., enter "07" for a wind from 70 degrees. When the wind speed equals 100 knots or more, add 50 to wind direction, e.g., enter "57" for a wind from 70 degrees when the speed is 100 knots or more. Enter "99" 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 12 for more examples.

Table 12. 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*). Enter the 10 meter forecast mean wind speed for the MVF valid period to the nearest knot, not to the nearest 5 knots, as expressed in the worded forecasts. Do not forecast 99 knots. If the wind speed is less than 10 knots, enter a zero in the tens digit place, e.g., enter "06" for 6 knots. For speeds of 100 knots or more, subtract 100 from the forecast speed and add 50 to the forecast direction. For example, given a forecast 110 knot wind from 270 degrees, enter “77” for wind direction and “10” for wind speed. See paragraph e. and Table 12 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*). Enter mean significant wave height for the MVF valid period in feet. If less than 10 feet, place a zero in the tens digit, e.g., enter "08" 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) enter "99" 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 will access verification statistics through the *Stats on Demand* feature of the NWS Verification Web Page. *Stats on Demand* will access an interactive database that will provide verification statistics customized to the user’s request. The user will request data for any marine weather element and the desired guidance product for one or more

a. months;

b. model cycle (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).

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

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

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

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

(5) A wind 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.

b. Wind Speed. The coded forecast to the nearest knot is verified against the mean of the five hourly wind speed observations during the MVF valid period. The observations used in verification may vary considerably in height and are corrected to the 10 meter standard forecast height by NDBC (Liu et al. 1979). Verification statistics are computed from the information contained in 7x7 contingency tables of forecasts versus observations. The wind speed categories are:

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.

c. Wind Direction. 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.

Verification statistics are computed from the information contained in 8x8 contingency tables of forecasts vs. observations. The categories are defined as 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.

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

1: Less than 3 feet,
2: 3 to 5 feet,
3: 6 to 8 feet,

4: 9 to 12 feet,
5: 13 to 16 feet,
6: 17 to 20 feet,
7: Greater than 20 feet.

3.2 Coastal Flood and Lakeshore Flood Warnings. WFOs with marine forecast responsibility perform coastal flood warning/lakeshore flood warning (CFW) verification manually.

3.2.1 Matching Warnings and Events. Treat each public forecast zone as a separate verification area. Therefore, count a warning covering three zones as three warned areas or three warnings. Only the following reportable events in a WFO's monthly *Storm Data* report verify a CFW.

a. Storm surge. The only storm surges verifying a CFW are (1) storm surges from extratropical cyclones and (2) storm surges resulting from a gradient induced between a tropical cyclone and strong high pressure.

b. Seiche.

c. High astronomical tide.

Treat Minor coastal or lakeshore flooding, such as nuisance flooding, as a non-event for verification purposes.

Record warnings and events in separate databases. All listings in the event database must meet warning criteria. Do not record multiple events for a single zone. Count one verified warning and one warned event whenever the event occurs in a warned zone. Count one unwarned event if the event occurs in a zone with no warning. Record one unverified warning for each warned zone that does not experience an event.

3.2.2 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. Count each extension in time of a zone already warned as a new warning if the issuance time of the extension preceded the time warning criteria were first met.

3.2.3 Lead Time. Compute a lead time for each public forecast zone that experiences a coastal or lakeshore flood event. Subtract the time of warning issuance from the time when the event first met warning criteria in the zone. Set negative values to zero. If a zone experiences an event meeting warning criteria with no warning 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.

3.2.4 Reports. The regional headquarters will report verification statistics to the OCWWS Marine and Coastal Weather Services Branch. Use the format in Table 13. All reports for the previous extra-tropical storm season, i.e., July through June, are due by the fifth working day of

August.

Table 13. Example of format for reporting CFW verification statistics.

Verification Statistics **Fiscal Year _________** **Date of Report ________** **Region __________**	**CFW**
Number of Warnings Issued	
Number of Verified Warnings	
Number of Unverified Warnings	
Number of Events	
Number of Warned Events	
Number of Unwarned Events	
Average Lead Time	
Probability of Detection (POD)	
False Alarm Ratio (FAR)	
Critical Success Index (CSI)	

3.3 Special Marine Warnings (SMW). The OCWWS Performance Branch operates and maintains the automated SMW verification program. Any SMW issued for a coastal or Great Lake marine zone, Lake Okeechobee, or Lake Pontchartrain is verified.

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

a. 3/4 inch or greater marine hail.

b. Marine thunderstorm wind, 34 knots or greater.

c. Waterspouts.

Warnings and events are recorded in separate databases. Whenever an event occurs in a warned marine zone, the following are recorded: one verified warning and one warned event. One unwarned event is recorded for each event that occurs in a zone with no warning. One unverified warning is counted for each warned zone that does not experience an event.

3.3.2 Quality Assurance Rules. In an attempt to reduce the impact of erroneous short-fused warnings on users and, at the same time, more accurately measure the quality of NWS warnings, the OCWWS Performance Branch has developed a set of rules stating how these short-fused warnings are archived.

a. Rule 1 – How Warnings are Entered into the Database. All data imported into the warning database are taken directly from the warning. No data are entered into the database from any information other than that represented by the bold-faced parts of the warning sample in Table 14. Based on this rule, products issued with the improper coding may or may not be imported into the database. Several examples appear on the NWS Verification Web Page.

Table 14. Special marine warning sample. The warning issuance and expiration date/times are taken from the VTEC line. The issuing WFO and warning type are checked for consistency with the top two lines of the warning. Inconsistent warnings are not counted for verification.

```
WHUS51 KWSH 010000
SMWWSH
DMZ001-003-005-010200-
/O.NEW.KWSH.MA.W.0010.050101T0000Z-050101T0200Z/

BULLETIN - EAS ACTIVATION REQUESTED
SPECIAL MARINE WARNING
NATIONAL WEATHER SERVICE SILVER SPRING MD
700 PM EST WED JAN 1 2001

THE NATIONAL WEATHER SERVICE IN SILVER SPRING HAS ISSUED A

* SPECIAL MARINE WARNING FOR...
  NORTHERN JEFFERSON BAY
  REAGAN POINT TO WASHINGTON LIGHTHOUSE
  KENNEDY HARBOR

* UNTIL 900 PM EST

* AT 700 PM EDT...A STORM SPOTTER REPORTED A WATERSPOUT ONE MILE
       OF NORTH OF REAGAN POINT...MOVING SLOWLY TOWARDS THE NORTHEAST.

A WATERSPOUT IS A TORNADO OVER WATER THAT CAN BE DANGEROUS AND EVEN
DEADLY. SMALL CRAFT CAN EASILY BE SWAMPED OR OVERTURNED BY A
WATERSPOUT. STAY AWAY FROM THEM AT ALL TIMES.

LAT...LON 3778 9752 3748 9752 3749 9724 3785 9724

$$
```

b. Rule 2 – Quality Assurance of Overlapping Warnings. This is valid for all warnings issued from January 1, 2002 to February 28, 2005. When two warnings for a given zone overlap in time, the portion of the earlier warning that overlaps the second warning is removed. The expiration time of the first warning is changed to one minute before the issuance time of the second warning. Several examples appear on the NWS Verification Web Page.

3.3.3 Lead Time. The lead time for each event is computed separately for each marine zone by subtracting the time of warning issuance from the time when warning criteria were first met in the zone. The time of warning issuance is taken from the VTEC line. 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.

3.3.4 Display of Verification Statistics. NWS employees access verification statistics through the *Stats on Demand* feature of the NWS Verification Web Page. *Stats on Demand* accesses an interactive database that provides verification statistics customized to the user's request. The user may request data by:

a. one or more dates (select beginning and ending date);

b. one or more marine zones, 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).

3.3.5 Backup Mode for Warnings. When a WFO goes into backup mode, warnings are still sorted by county, so all warnings issued by the backup office are attributed to the primary WFO.

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

4.1 Flash Flood Warnings. The OCWWS Performance Branch is responsible for the operation and maintenance of the automated FFW verification program.

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

For verification purposes, multiple flash flood events in the same county separated by less than 30 minutes are considered duplicates; therefore, only the first entry is made to the event database. 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. 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 still appears in the publication *Storm Data*.

Warnings and events are recorded in separate databases. Whenever an event occurs in a warned county, the following are recorded: one verified warning and one warned event. One unwarned event is recorded for each event that occurs in a county with no warning. One unverified warning is counted for each warned county that does not experience an event.

4.1.2 Quality Assurance Rules. In an attempt to reduce the impact of erroneous short-fused warnings on users and, at the same time, more accurately measure the quality of NWS warnings, the OCWWS Performance Branch has developed a set of rules stating how these short-fused warnings are archived.

a. Rule 1 – How Warnings are Entered into the Database. All data imported into the warning database are taken directly from the warning. No data are entered into the database from any information other than that represented by the bold-faced parts of the warning sample in Table 15. Based on this rule, products issued with the improper coding may or may not be imported into the database. Several examples appear on the NWS Verification Web Page.

b. Rule 2 – Quality Assurance of Overlapping Warnings. This is valid for all flash flood warnings issued on or after January 1, 2002, and will be discontinued for all FFWs issued after the implementation of VTEC in the FFW product. When two warnings for a given county overlap in time, the portion of the earlier warning that overlaps the second warning is removed. The expiration time of the first warning is changed to one minute before the issuance time of the second warning. Several examples appear on the NWS Verification Web Page.

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 “date and time issuance” line of the warning (see Table 15). Once VTEC is implemented in the FFW product, the issuance time will come be taken from the VTEC line in the FFW. Negative values are converted 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.

Table 15. Flash flood warning sample.

```
WGUS51 KWSH 010000               <------ WMO PRODUCT ID
FFWWSH                           <------ WARNING TYPE and WFO
DCC001-003-005-010200-           <------ COUNTY and STATE WARNED

BULLETIN - EAS ACTIVATION REQUESTED
FLASH FLOOD WARNING
NATIONAL WEATHER SERVICE SILVER SPRING MD
700 PM EST WED JAN 1 2001        <------ DATE AND ISSUANCE TIME

THE NATIONAL WEATHER SERVICE IN SILVER SPRING HAS ISSUED A

* FLASH FLOOD WARNING FOR...
  WASHINGTON COUNTY IN THE DISTRICT OF COLUMBIA
  JEFFERSON COUNTY IN THE DISTRICT OF COLUMBIA
  REAGAN COUNTY IN THE DISTRICT OF COLUMBIA

* UNTIL 900 PM EST               <------ EXPIRATION TIME

* AT 700 PM EST...DOPPLER RADAR INDICATED A THUNDERSTORM PRODUCING HEAVY
     RAINS 2 MILES WEST OF ADAMS MORGAN...MOVING EAST AT 15 MPH.

EXCESSIVE RUNOFF WILL CAUSE FLOODING OF LOW LYING AREAS AND
INTERSECTIONS. POOR DRAINAGE AREAS ARE MOST AT RISK. DO NOT DRIVE
THROUGH FLOODED ROADS.


LAT...LON 3778 9752 3748 9752 3749 9724 3785 9724

$$
```

4.1.4 Display of Verification Statistics. NWS employees access FFW verification statistics through the *Stats on Demand* feature of the NWS Verification Web Page. *Stats on Demand* accesses an interactive database that provides verification statistics customized to the user's request. The user may request data by:

a. one or more dates (select beginning and ending date);

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

4.1.5 Backup Mode for Warnings. When a WFO goes into backup mode, FFWs are still sorted by county, so all FFWs issued by the backup office are attributed to the primary WFO.

4.2 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. Monthly verification statistics are automatically sent from the RFCs to the OCWWS Performance Branch.

NWS employees access verification statistics through the *Stats on Demand* feature of the NWS Verification Web Page. *Stats on Demand* accesses an interactive database and generates 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 web.

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. 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 Web Page.

6. Aviation Verification Procedures.

6.1 Legacy TAF Verification.

6.1.1 Introduction. The National Weather Service (NWS) maintains a long-term record of terminal aerodrome forecast (TAF) verification statistics for ceiling and visibility. Quality controlled data are archived in a centralized database and are available to NWS employees through query, using the *Stats on Demand* feature of the NWS Verification Web Page. A list of verification sites, known as the legacy network, appears on this Web site, and data are available for most of these locations from October 1997 to September 2004. Each Weather Forecast Office (WFO) has at least one legacy verification site in its area of forecast responsibility, and a list of these sites appears on the NWS Verification Web Page. The legacy system no longer imports new data. For information about verification of recent TAFs and the modernized TAF verification system, see section 6.2.

The NWS Verification 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 and generates verification statistics customized to the user's request. The user may request data for ceiling and visibility and the desired MOS guidance product for one or more:

a. months,

b. TAF beginning times,

c. projections,

d. verification sites (single site, multiple sites, regional, or national).

6.1.2 Elements. Only regularly scheduled TAFs beginning at 0600 and 1800 UTC are included. Projections for TAF elements are defined as the number of hours elapsed since the beginning time of the TAF. The verification software evaluates the prevailing portion of the TAF and does not recognize temporary change (TEMPO) groups and probability (PROB) groups. Amended forecasts are not included.

a. Ceiling Height.

(1) Projections: 3, 6, 9, and 15 hours.

(2) TAF: The TAF ceiling height at each verifying hour was recorded in hundreds of feet above ground level (AGL).

(3) Guidance: The MOS guidance (GFS and NGM) forecasts at each verifying hour were recorded. The MOS cycle that initialized 6 hours before the TAF beginning time was used.

(4) Observations: From METAR, ceiling height at each verifying hour was recorded in hundreds of feet AGL.

(5) Verification Statistics: NWS employees may run *Stats on Demand* to generate and display contingency tables and verification statistics for the TAF and one of the MOS guidance products. The contingency tables consist of TAF/MOS guidance data versus observations using the following categories:

1: Less than 500 feet,
2: 500 to 900 feet,
3: 1000 to 3000 feet,
4: Greater than 3000 feet (includes cases with no ceiling).

A second breakdown of categories is also available:

A: Less than 200 feet,
B: 200 feet or above (includes cases with no ceiling).

b. Visibility

(1) Projections: 3, 6, 9, and 15 hours.

(2) TAF: The TAF visibility at each verifying hour was recorded in statute

miles and fractions thereof. TAF visibilities above 6 statute miles were recorded as "7."

(3) MOS Guidance: The MOS guidance (GFS and NGM) forecasts at each verifying hour were recorded. The MOS cycle that initialized 6 hours before the TAF beginning time was used.

(4) Observations: From METAR, visibility at each verifying hour was recorded in statute miles and fractions thereof.

(5) Verification Statistics: NWS employees may run *Stats on Demand* to generate and display contingency tables and verification statistics for the TAF and one of the MOS guidance products. The contingency tables consist of TAF/MOS guidance data versus observations using the following categories:

1: Less than 1 statute mile,
2: 1 through less than 3 statute miles,
3: 3 through 5 statute miles,
4: Greater than 5 statute miles.

A second breakdown of categories is also available:

A: Less than or equal to 1/4 statute mile,
B: Greater than 1/4 statute mile.

6.2 Modernized TAF Verification Program. This *Stats on Demand* program is the official NWS TAF verification tool. TAFs are evaluated twelve times per hour or 288 times for an entire 24-hour TAF—at the end of every 5-minute interval whose clock time to the nearest minute ends in "0" and "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 until a new METAR or SPECI is reported.

6.2.1 Verification Sites. All terminals for which the NWS issues TAFs may be verified. A list of all TAF verification sites appears on the NWS Verification Web Page.

6.2.2 Data Input. All data are automatically collected from operational products by OCWWS. 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 Verification Web Page. Guidance data come from the alphanumeric MOS products derived from the GFS model and NGM. The Local AWIPS MOS Program (LAMP) product, derived from the NGM 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.

6.2.3 TAF Verification Reports. NWS employees access verification statistics through the *Stats on Demand* feature of the NWS Verification Web Page. *Stats on Demand* accesses an interactive database and generates 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., MOS, LAMP, persistence) for one or more:

a. months,

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

c. projection period groups (see section 6.2.4),

d. verification sites (single site, multiple sites, WFO forecast area, NWS Region, or national). When a single WFO forecast area or a subset of it is selected, a forecaster may use his/her private password to request verification statistics from *Stats on Demand* that include only the forecasts made by that forecaster. The personalized password will protect the privacy of each forecaster and keep individualized verification statistics confidential.

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. Forecast categories for each element are defined in section 6.2.5.

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

a. greater than zero to 3 hours,

b. greater than 3 to 6 hours,

c. greater than 6 to 9 hours,

d. greater than 9 to 12 hours,

e. greater than 12 to 24 hours.

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:

a. greater than zero to 3 hours,

b. greater than 3 to 6 hours.

6.2.5 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. 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. Sometimes categories are combined.

   1: Less than 200 feet,
   2: 200 to 400 feet,
   3: 500 to 900 feet,
   4: 1000 to 1900 feet,
   5: 2000 to 3000 feet,
   6: Greater than 3000 feet (includes cases with no ceiling).

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

   1: Ceiling less than *x*,
   2: 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. From these categories, contingency tables of forecasts versus observations and guidance versus observations are prepared, and verification statistics are computed. Sometimes categories are combined.

   1: Less than ½ statute mile,

2: ½ to less than 1 statute mile,
3: 1 through less than 2 statute miles,
4: 2 through less than 3 statute miles,
5: 3 through 5 statute miles,
6: Greater than 5 statute miles.

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

1: Visibility less than *y,*
2: 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 16. The categories for ceiling and visibility are then combined by taking the lower category of the two. This is the flight category. From these categories, contingency tables of forecasts versus observations and guidance versus observations are prepared, and verification statistics are computed. Sometimes categories are combined.

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 16. 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. Wind direction is not verified whenever (1) the observed speed is less than 6 knots or (2) 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 16 knots,
2: 16 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. Weather Type. Each of the following weather types is verified separately in two 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 only forecasts weather type (1) (fog

types). NGM MOS only forecasts weather types (1) and (2). NGM LAMP only forecasts weather types (1), (2), (3), (4), and (6).

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

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

(3) Freezing precipitation—freezing rain (FZRA), freezing drizzle (FZDZ),

(4) 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),

(5) Thunderstorms (TS), including funnel clouds (FC) and tornadoes/waterspout (+FC). Some observation stations do not report thunderstorms. These METARs use the TSNO remark. Thunderstorms in the TAF may not be verified under these conditions.

(6) Hail (1/4 inch or greater diameter) (GR) and showers of GR (SHGR),

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

(8) Haze (HZ) and smoke (FU),

(9) 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).

(10) Blowing spray (BLPY),

(11) Blowing snow (BLSN), drifting snow (DRSN),

(12) Volcanic ash (VA), and

(13) Squalls (SQ).

6.2.6 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 paragraph a. Another “type” of forecast, called the operational impact forecast, is defined in paragraph b. 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 is explained, respectively, in paragraphs c and d. The following terms will be repeated several times in paragraphs a through d and are defined:

(1) Change. For ceiling and visibility, change is defined as category change. Categories for these elements are defined in section 6.2.5. Each of the thirteen weather types is a binary variable, and change is defined as the starting or stopping of that weather type. Precipitation intensities are ignored. For wind direction, change is defined (a) as a 40-degree or greater wind shift between successive observations, considering only 6-knot or greater observations or (b) by a variable wind remark. For sustained wind speed, change is defined as at least an 8-knot increase or decrease between successive observations. For wind gusts, change is defined as (a) at least a 10-knot increase or decrease between successive observations or (b) when successive observations change from the existence of gusts to no gusts or vice versa.

(2) Hit. For ceiling and visibility, a forecast hit is defined as the forecast category equaling the observation category, and categories are defined in section 6.2.5. For each of the thirteen weather types, a hit occurs when the forecast and observation agree on the occurrence or non-occurrence of that weather type. For sustained wind speed, a hit occurs whenever the absolute error is less than 8 knots. 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*.

(3) Less [More] in Error. When comparing two forecast types (i.e., prevailing and TEMPO, prevailing and PROB) 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 6.2.5) . 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. All thirteen weather types are binary variables, so the term “less [more] in error” is not used when referring to any of them.

(4) More [Less] Favorable Flight Conditions. When comparing two forecasts types (i.e., prevailing and TEMPO, prevailing and PROB) 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 6.2.5. For each of the thirteen weather type forecasts (each is a binary variable), the more [less] favorable flight conditions are defined as the negative [positive] forecast of the event. 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. Wind direction forecasts are not compared in this manner.

a. 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 6.2.5, 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.

b. Operational Impact (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). (3) The prevailing forecast is in effect simultaneously with a 30% or 40% 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:

(1) The OIF is undefined for wind direction.

(2) 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.

(3) 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.

(4) 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.

(a) *First step—the variability test.* The legitimacy of the TEMPO forecast is first evaluated by a variability test at the end of every 5-minute interval of the TAF. If the observation database changed twice or more ±90 minutes from the end-point of the 5-minute interval, then the TEMPO forecast passes the variability test for that 5-minute interval. Note: This test just measures condition variability—it does not measure forecast correctness.

(b) *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 weather type.

If the TEMPO forecast passes the variability test for a given 5-minute interval, then the OIF for that interval is defined as (1) the forecast with the smallest categorical error for ceiling and visibility; (2) the smallest error for wind speed and wind gusts; or (3) no error for weather type.

(5) 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 16. 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 6.2.5, 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.

c. <u>TEMPO Forecast</u>. 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:

(1) <u>Number of hours</u>. This is the total number of hours (the number of 5-minute intervals divided by 12) that TEMPO groups were valid for the user-specified element. Data are given to the nearest hour.

(2) Justified TEMPO (%). This is the percentage of 5-minute intervals within the TEMPO groups that passed the aforementioned OIF variability test for the user-specified element and is re-stated: For each 5-minute interval inside a TEMPO group, if the observation database changed twice or more ±90 minutes of the end-point of that 5-minute interval, then the TEMPO forecast for that 5-minute interval is justified. This statistic just 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 % time that the 0800-1200 UTC TEMPO group was justified was 10/48 = 21%. After 0850 UTC, none of the 5-minute intervals were "justified"(no changes or only one change occurred ±90 minutes of the end time of each 5-minute interval). Hence the justification test passed between 0800 and 0850 UTC and failed after 0850 UTC, making the numerator 10. The denominator is 48 (12 possible changes per hour times 4 hours).*

(3) 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. TEMPO Hit (%): 40.*

(4) Justified TEMPO–Improved the TAF (%). Considering only the 5-minute intervals when the TEMPO forecast was justified, this is the percentage of time that the TEMPO forecast was not a hit; however the TEMPO forecast was less in error than the prevailing forecast. Since each of the thirteen weather types are binary variables and can only hit or miss, they are not evaluated with this statistic. Ideally, 10 to 49 percent of these TEMPO cases 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. TEMPO improved TAF (%): 40.

(5) 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.*

(6) 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.*

(7) 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 ultimately do 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.*

d. PROB Forecast. The PROB forecast is evaluated at the end of every 5-minute interval that the PROB forecast is valid 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:

(1) 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.

(2) 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 is sufficient 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.*

(3) 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.*

(4) 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.

(5) 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.*

(6) 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 PROB 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 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.*

(7) 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.*

6.3 Aviation Verify (Oxnard) TAF Verification Program. This formerly transitional program is now optional throughout the NWS. However, WFOs may continue running it to provide continuity in verification results. Every 5-minute segment of every scheduled TAF is verified. The use of TEMPO and PROB groups in TAFs means two separate forecasts may be valid for a given terminal at the same time, the prevailing conditions and the TEMPO or PROB conditions. In the legacy verification program, just the prevailing conditions are verified, even though the TEMPO or PROB forecast conditions often have a larger impact on operations and flight planning. Aviation Verify also provides some tools to help the user evaluate TEMPO and PROB groups.

Paragraph 6.3 is out-dated: */Stats-on-Demand/* is the official verification tool (as noted in paragraph 6.2), including for individual forecasters. WFOs may continue to use the PC-based Aviation Verify (Oxnard) TAF Verification program to provide continuity in verification results, but are no longer required to do so.

6.3.1 Verification Sites. All terminals for which a WFO issues TAFs are verified.

6.3.2 Data Collection. At WFOs and the regional headquarters, all data used by Aviation Verify is collected locally via AWIPS. Ceilings, visibilities, wind direction, and wind speed data are collected from all scheduled TAFs, METARs, SPECIs, the NGM MOS, and the GFS MOS. Southern Region Headquarters ingests and temporarily stores for the current month all raw data for the entire Nation.

6.3.3 Data Transmission. At the beginning of each month, raw data from the previous month are automatically transmitted from Southern Region Headquarters to the OCWWS Aviation Services Branch.

6.3.4 Reports.

a. WFO Results. Verification statistics for each forecaster or the entire WFO may be computed for a designated period.

b. Regional Results. Using the raw data archived at each regional headquarters office, verification statistics for the NWS region or any subset of the region may be computed for a designated period using the program RAMVer.

c. National Results. Using the data collected at OCWWS, verification statistics may be computed for the entire nation or any desired subset of the nation for a designated period using the program RAMVer.

6.4 Aviation Weather Center (AWC) Verification Procedures.

6.4.1 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 a new software system which 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.

6.4.2 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 Web sites:

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. Temperature and precipitation for the following forecasts are verified using first-order (CLIMAT international exchange) stations in the continental United States:

a. 6-10 day,

b. Week 2 (8-14 day),

c. monthly (issued with a 0.5 month lead time),

d. seasonal (0.5 to 12.5 month lead time).

The number of stations used in the verification varies between 60 and 100, depending on variable and time period. Temperature and precipitation for the extended lead seasonal forecasts (1.5 to 12.5 month lead time) are verified using climate division data from the National Climatic Data Center. Because these data only become available after 2 to 3 months, verification of these forecasts is delayed. A version of the Heidke Skill score (described in appendix A, section 2.c) is computed for climate forecast 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 Tuesday for the 3- to 14-day forecast period, i.e., 1200 UTC Friday (Day 3) until 1200 UTC on the Tuesday two weeks after the forecast is issued (Day 14). The hazard assessment may be updated anytime before 1200 UTC Friday, Day 3; all issuances (scheduled and unscheduled updates) 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 SPC's 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 17. 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 the contingency table). For example, a heavy precipitation hazard was forecast for a particular grid point from November 17 thru 19, and 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.

CPC also computes two additional scores not included in appendix A, section 3:

a. Hit Rate Score.

$$HR = \frac{A + X}{A + B + C + X}$$

b. Bias.

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

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.

Beasley, R.A., 1995: Automation of Field Operations and Services (AFOS)-era forecast verification. NOAA Techniques Development Laboratory Computer Program NWS TDL CP 95-2, National Weather Service, NOAA, U.S. Department of Commerce, 50 pp.

Dagostaro, V.J., 1985: The national AFOS-era verification data processing system. TDL Office Note 85-9, National Weather Service, NOAA, U.S. Department of Commerce, 47 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 forecast/observation contingency table is often developed to summarize all variables by category. Table A-1 contains a generalized contingency table with k mutually exclusive and exhaustive categories. Each element of the table, A_{ij}, gives the number of times the observation was in the ith category and the forecast was in the jth category. The row and column totals, respectively R_i and C_i, are often called the marginal totals of the contingency table.

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_t	C_2	...	C_k	N

Various scores can be computed from the elements in a contingency table such as:

2.1 Percent Correct *(PC)* is the percentage of time a correct forecast was made *(i=j)* regardless of the category.

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

2.2 Bias by Category (BIAS) measures the tendency to overforecast (*BIAS* greater than 1) or underforecast (*BIAS* less than 1) a particular category, *i*. In Table A-1, *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*.

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

2.4 False Alarm Ratio (FAR). An *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.

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

2.5 Critical Success Index (CSI). 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.

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

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

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

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

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.* Heidke does not give partial credit for “near hits” in situations with 3 or more categories.

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

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

A perfect Heidke skill score is one. Zero is indicative of no skill, and a negative score indicates skill worse than random forecasts.

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 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. Neither Heidke nor Peirce give partial credit for “near hits” in situations with 3 or more categories.

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

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

2.9 Equatable Skill Scores (ESS).

2.9.1 Subjective Explanation. Equitable skill scores are often used to evaluate multi-category forecasts. Gandin and Murphy (1992) introduced a score that uses weighted event probabilities of each category to attain consistency or equitability (Gandin and Murphy 1992). Gerrity (1992) constructed a subset of the Gandin and Murphy score for application to ordinal multi-categorical event forecasts. These scores are zero when produced from a set of randomly generated forecasts; therefore, the score cannot be "gamed" by a clever forecaster. A perfect set of forecasts results in values equal to one, and a negative score results from a set of forecasts with less skill than random forecasts. 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.

These non-linear properties discourage forecaster "hedging" by penalizing forecasters who frequently forecast the most climatologically likely events and rewarding forecasters who correctly forecast rare events. The property of giving large rewards for correct forecasts of rare events may make the score volatile, especially with small sample sizes. In other words, if a particular event occurs on a rare basis, the ESS may increase substantially due to just one additional correct forecast of that rare event. 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 in the contingency table. With ceiling and visibility, the rarest events tend to be the lowest categories in the contingency table. 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. Hence, the ESS is not the ideal score for data requests that include relatively small geographic areas and/or relatively short periods of time. Whenever the ESS is used, the delta values should be checked because they warn of potential volatility in the score. 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.

Prior to May 13, 2003, the NWS marine verification program computed the ESS from a multi-year national climatology, regardless of the area or time frame specified in the *Stats on Demand* data request. The ESS is now computed from the "climatology" of the geographic area and time period of the data request. The impact of this change on previously developed NWS milestones is expected to be minimal, because the baselines of scores that were used to develop these milestones came from data sets that included the entire Nation and all months of the year for several years. However, these national, full year milestones are only relevant when applied to national data from all months of the year.

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) introduced 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 circular. Wind speed and ceiling height are examples of ordinal, non-circular variables. Wind direction is an example of an ordinal, circular variable for which the Gerrity solution is not appropriate, because as an 8-category variable, wind direction can only "miss" by up to four categories (a non-circular variable could miss by up to seven categories). With nominal elements (i.e., order does not matter), the Gerrity equations are not appropriate due to the graduated reward-penalty system. Note, however, that nominal elements are rare in meteorology. Gerrity (1992) 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:

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

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

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

Burroughs (1993), appendix B, section n, applies these general equations for populating the **S** matrix to specific *k*-category marine elements. Burroughs (2002) illustrates these applications with examples. For more detail, the user is encouraged to consult the other references.

In the past, the NWS marine verification program computed the Gerrity ESS from a multi-year climatology that included available buoys and CMAN stations in the offshore and coastal water areas for which the agency issues forecasts. This national multi-year climatology was used to build a static **S** matrix which was used for all computations of the score. The multi-year national record helped minimize random fluctuations in the score and proved very effective in a system that emphasized national scores. A static climatology, however, is less effective in a *Stats on Demand* system which encourages the user to request data from geographic and temporal subsets of the national multi-year database. Sometimes these subsets can be as small as one buoy or CMAN station for a single month. Through a performance measure tutorial, Burroughs (2002) demonstrated that when the static **S** matrix was used to compute the Gerrity score on subsets of wind speed data with higher frequencies of strong winds and lower frequencies of light winds than the national dataset, the upper bound of the score for a “perfect” set of forecasts (100% correct categorically) was well above one (2.66 in Burroughs’ example). Conversely, when the static **S** matrix was used to compute the Gerrity score on subsets of wind speed data with lower frequencies of strong winds and higher frequencies of light winds, the upper bound of the score for a “perfect” set of forecasts was well below one (0.63 in Burroughs’ example). Burroughs addressed this problem by normalizing all scores to a maximum value equal to one.

An alternative solution to this problem would be to use a more appropriate climatology “tailored” to the data request by building a new **S** matrix every time the score is computed. Such an approach would incorporate a subset of the full multi-year climatological data set to develop the **S** matrix, using only the months of the year and geography consistent with the data request. This process would be more computationally intensive than before, but the **S** matrix would be more climatologically appropriate for each individual data request.

Livezey (2003) argues for a simpler approach—compute the **S** matrix directly from the sample of the individual data request. This would result in an **S** matrix that contains data only from the geography and time period of the data request. He argues that an **S** matrix built from long climatological records provides biased results when applied to verification statistics because climate change may have occurred during or since the period of climatological record. Given Livezey's point about climate change and the need in *Stats on Demand* to keep computations as simple as possible (the user works interactively in the system and waits for the results of his/her request), the Livezey recommendation was implemented in the NWS marine verification program on May 13, 2003. The **S** matrix is now computed directly from the sample of the *Stats on Demand* data request. This new methodology has one major shortcoming. Requests for verification data from relatively small 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 address 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 presented 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 hand. Divide 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*).

Given this procedure change in building the **S** matrix, the user should not compare scores that were computed using the old methodology with scores now computed from the NWS verification web site, except when the latter scores have been computed from national data sets that include all months of the year. The impact of this change on previously developed NWS milestones is expected to be minimal because the baselines of scores that were used to develop these milestones came from national data sets that included all months of the year for several years. As long as the milestones are applied to national data that include all months of the year, the effect of the change should be minimal.

3. Specialized Contingency Table. 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. Specialized Contingency Table

		Forecasts	
		Yes	No
Events	Yes	*A*	*B*
	No	*C*	*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).

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 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.

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

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 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.

Overforecasting 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. <u>Scores Computed for Specific Forecast Elements</u>. 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 <u>Temperature, Wind Speed and Direction, and Wave Height</u>. Scores frequently computed for forecasts of temperature, wind speed and direction, and wave height include:

a. <u>Mean Error *(ME)*</u> 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. <u>Mean Absolute Error *(MAE)*</u> 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 the 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, 0.02, 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 the Event *(RF)* is the fraction of the time the event occurred.

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

g. Reliability, a measure of bias, compares the average forecast of the event with the relative frequency of the event. The reliability may be determined overall or by forecast interval, e.g., 10 percent PoP intervals.

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

where, N is the total number of events or the number of events in the interval. If the average forecast of the event is larger (smaller) than the relative frequency of the event, the event 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). Based on the contingency table presented in section 3 of this manual,

$$POD(N) = \frac{X}{X + C}$$

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, 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 Terms

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.

Change - This term is used in defining the Operational Impact Forecast