NATIONAL WEATHER SERVICE INSTRUCTION 10-1601

November 9, 2015

Operations and Services

Performance, NWSPD 10-16

VERIFICATION

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

OPR: W/COO11 (B. McNulty) **Certified by:** W/COO1 (C. Woods)

Type of Issuance: Routine

SUMMARY OF REVISIONS: This directive supersedes National Weather Service Instruction (NWSI) 10-1601, dated September 28, 2011. The following changes have been made to the directive:

1. Updated template used to format the directive.
2. Removed old Appendix A as the content is now available on the Performance and Evaluation Branch website (https://verification.nws.noaa.gov/) and non-directive in nature.
3. Updated office names to reflect the recent reorganization of the NWS headquarters.
4. Moved background material on verification science into a new Appendix A.
5. Revised the main body to include only directive information about verification in the NWS.

/Signed/ 27 October 2015

John D. Murphy Date

Chief Operating Officer

Title of Directive

Table of Contents **Page**

1 Introduction

The National Weather Service (NWS) routinely reviews the quality and distribution of its warnings and forecasts, especially after catastrophic events result in the loss of human life and property. Such events can impact local economies and the national economy. Applying verification to the forecast process should produce actions leading toward improved service.

2 Mission Connection

Verification and impact assessments of forecasts drive the continuous improvement cycle needed to provide the nation with the best possible forecasts. Strong performance and evaluation feedback strengthens NWS forecast skills, and leads to improvement of weather products and services.

3 Forecast Verification

Verification is the process of matching warnings and forecasts with corresponding weather observations to assess the goodness of those warnings and forecasts. The observations vary from networks of instrument systems, which measure weather elements, to human spotter reports, where someone reports the details of a weather occurrence, such as a tornado or flash flood he/she just experienced or investigated. Appendix A is a discussion of the science of verification.

4 Verification Information and the Evaluation of Forecaster Performance

Verification scores are not used to establish criteria for rating the forecasting and warning 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 full 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 should be kept confidential.

4.1 Use of Verification by Managers

An explanation of NWS verification systems and their applications is available on the Performance Management website: https://verification.nws.noaa.gov/. 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 that are 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.

4.2 Verification Uses Specific to Aviation Services

To ensure forecaster privacy, only the WFO management team and the aviation focal point may access TAF verification statistics sorted by each individual forecaster at the WFO. This privacy is accomplished automatically through a system of usernames and passwords.

Each forecaster may request aviation TAF verification statistics for their WFO, or a subset of the WFO, that:

- Include only the TAFs issued by that forecaster.
- Reflect the verification statistics for all TAFs issued by the WFO.

APPENDIX A - Verification Science

1 Reasons to Verify

The Joint Working Group on Forecast Verification Research (JWGFVR) website www.cawcr.gov.au/projects/verification addresses the question, "why verify?"

"A forecast is like an experiment – given a set of conditions, you make a hypothesis that a certain outcome will occur. You wouldn't consider an experiment to be complete until you determined its outcome. In the same way, you shouldn't consider a forecast experiment to be complete until you find out whether the forecast was successful."

"The three most important reasons to verify forecasts are:

- To monitor forecast quality – how accurate are the forecasts and how are they improving over time?
- To improve forecast quality – the first step toward getting better is discovering what you're doing wrong.
- To compare forecast quality of different forecast systems – to what extent does one system give better forecasts than another and in what ways is that system better?"

2 Forecast Goodness

Weather forecasts have high quality if they predict the observed conditions well. Murphy (1993) broadened the topic of quality to forecast goodness, of which there are three types:

- Type I: Consistency is the degree to which the forecaster's judgments from his/her knowledge base correspond to the actual warning or forecast.
- Type II: Quality is the degree to which the warning or forecast corresponds to what actually happened by comparing forecasts to a corresponding set of observations.
- Type III: Value is the incremental economic or other benefits realized by decision makers through the use of warnings and forecasts.

The JWGFVR Website provides an illustration of the difference between quality and value:

Forecast quality is not the same as forecast value. A forecast has high quality if it predicts the observed conditions well according to some objective or subjective criteria. It has value if it helps the user to make a better decision.

An example of a forecast with high quality but of little value is a forecast of clear skies over the Sahara Desert during the dry season.

When the cost of a missed event is high, the deliberate over-forecasting of a rare event may be justified, even though a large number of false alarms may result. An example of such a circumstance is the occurrence of fog at airports.

3 Forecast Quality

The JWGFVR Website www.cawcr.gov.au/projects/verification defines nine aspects of forecast quality (Murphy called them attributes):

- Bias: the correspondence between the mean forecast and the mean observation.

- Accuracy: the level of agreement between the forecast and the truth (as represented by observations). The difference between the forecast and the observation is the error. The lower the errors, the greater the accuracy.

- Skill: the relative accuracy of the forecast over some reference forecast. The reference forecast is generally an unskilled forecast such as random chance, persistence, or climatology. Persistence is defined as the most recent observation at forecast time and implies no forecasted change in condition. Skill refers to the increase in accuracy due purely to the "smarts" of the forecast system. Weather forecasts may be more accurate simply because the weather is easier to forecast— skill takes this into account.

- Reliability: the average agreement between the forecast values and the observed values. If all forecasts are considered together, the overall reliability is the same as the bias. If the forecasts are stratified into different ranges or categories, then the reliability is the same as the conditional bias, i.e., it has a different value for each category.

- Association: the strength of the linear relationship between the forecasts and observations (for example, the correlation coefficient measures this linear relationship).

- Resolution: the ability of the forecast to sort or resolve the set of events into subsets with different frequency distributions. This means that the distribution of outcomes when "A" was forecast is different from the distribution of outcomes when "B" is forecast. Even if the forecasts are wrong, the forecast system has resolution if it can successfully separate one type of outcome from another.

- Sharpness: the tendency of the forecast to predict extreme values. To use a counter-example, a forecast of climatology has no sharpness. Sharpness is a property of the forecast only, and like resolution, a forecast can have this attribute even if it is wrong (in this case it would have poor reliability).

- Discrimination: ability of the forecast to discriminate among observations, that is, to have a higher prediction frequency for an outcome whenever that outcome occurs.

- Uncertainty: the variability of the observations. Greater uncertainty is typically related to more difficult forecasts.

4 Skill

Weather forecast verification has traditionally focused on accuracy and skill. Skill scores are more helpful than accuracy in assessing forecast quality because skill scores subtract the effects of persistence, the climatological mean, or random chance from the forecasts. Sometimes forecasts based largely upon these parameters can appear to be good, especially in locations where persistence or the climatic mean are very prevalent, but the skill of such forecasts is often quite low. For example, if "no wind gusts or gusts less than 20 knots" occur at a given location 80 percent of the time, a set of forecasts that always predicts these conditions might appear to be skillful, i.e., the long-term percentage of correct forecasts would be 80. However, in this example the forecasts predict the most commonly observed low wind speeds all the time, thereby demonstrating no improvement over climatology or random chance, which is the definition of zero skill. Such a forecast provides no one with any advance information on the potential for dangerous or damaging winds.

5 Forecast Value

The measurement of forecast quality is essential, but quality measurements only reflect part of their overall contribution to society. Another important aspect of "forecast goodness" is the value they provide to users. Forecast value can be described in various ways. As defined in section 2.2, it is the incremental economic or other benefits realized by decision makers through the use of the warnings and forecasts. Jolliffe and Stephenson (2003) define the value of a forecast system as "the reduction in mean expense relative to the reduction that would be obtained by having access to perfect forecasts" (page 168). Lazo et al. (2009) described value in economic terms by indicating that it "represents the trade-offs people are willing to make to receive this information relative to other information, goods, or services" (page 786).

The NWS intends to add forecast value to fulfill its mission to save lives and enhance the national economy by providing superior Impact-Based Decision Support Services (IDSS). The difficulty is determining how to effectively measure this value. Subjective determination of the value of NWS IDSS can be partially achieved through user feedback (e.g., satisfaction surveys). However, obtaining meaningful and objective measurements of forecast value are more difficult.

The NWS is in the early stages of pursuing objective forecast value information. As a first step, more specific information and feedback from core partners and other users will be necessary to determine answers to such basic questions as:

What makes an event high impact?
What forecast elements are most important to operations?
What are critical weather, water, or climate element thresholds that trigger actions? Do specific NWS products or services trigger decision-maker actions?

As the NWS improves its understanding of societal impacts in general, and acquires knowledge about weather-related impacts on its core partners, impact-based verification should be developed to measure the true value of NWS services.

6 References

Jolliffe, I. T., & Stephenson, D. B. (2003). Forecast verification - a practitioner's guide in atmospheric science. Hoboken, NJ: John Wiley & Sons.

Lazo, J. K., Morss, R. E., & Demuth, J.L. (2009). 300 billion served: Sources, perceptions, uses, and values of weather forecasts. Bulletin of the American Meteorological Society, 90, 785–798.

Murphy, A. H. (1993). What is a good forecast? An essay on the nature of goodness in weather forecasting. Weather and Forecasting, 8, 281–293.

WMO (2010). Joint Working Group on Forecast Verification Research (JWGFVR) Website: www.cawcr.gov.au/projects/verification