NOAA Technical Memorandum NWS T&EL-15

PROGRESS AND RESULTS OF FUNCTIONAL TESTING (SUPPLEMENT TO NOAA TECHNICAL MEMORANDUM NWS T&EL-12)

Sterling, Va.
April 1977

noaa NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION / National Weather Service

NOAA TECHNICAL MEMORANDA

National Weather Service, Test and Evaluation Laboratory Series

The Test and Evaluation Laboratory (T&EL) of the Office of Systems Development conducts experiments and tests to determine the suitability of newly developed equipment, instruments, and systems for field implementation by the National Weather Service (NWS). Results of tests and evaluations are provided to the other operating offices of the NWS (Office of Meteorological Operations, Office of Hydrology, and National Meteorological Center) to guide field implementation activities, and to the Equipment Development Laboratory if further development is deemed necessary.

NOAA Technical Memoranda in the NWS T&EL series facilitate prompt distribution of scientific and technical material by staff members, cooperators, and contractors. Information presented in this series may be preliminary in nature and may be published formally elsewhere at a later date. Publications 1 through 9 are in the former series, ESSA Technical Memoranda, Weather Bureau Technical Memoranda (WBTM). Beginning with 10, publications are now part of the series, NOAA Technical Memoranda NWS.

Publications listed below are available from the National Technical Information Service (NTIS), U. S. Department of Commerce, Sills Bldg., 5285 Port Royal Road, Springfield, Va. 22151. Price: $3.00 paper copy; $1.45 microfiche. Order by accession number, when given, in parentheses.

ESSA Technical Memoranda

WBTM T&EL 1 Final Report - Test and Evaluation of Berkeley Automatic Station. James E. Morris, January 1967. (Not available)

WBTM T&EL 2 Final Report - Test and Evaluation of the Weather Bureau Radar-Telephone Transmission System (WM/RATTS-65). Robert E. Johnson, September 1967. (PB-176-532)

WBTM T&EL 3 Final Report - Test and Evaluation of the Mortarboard Psychrometer. Walter E. Hoehne and Roger A. Tucker, January 1968. (PB-177-689)

WBTM T&EL 4 Final Report - Test and Evaluation of a Facsimile Bandwidth Compression Technique. April 1968. (Not available)

WBTM T&EL 5 Final Report - Test and Evaluation of the Video Integrator and Processor. R. C. Strickler, April 1968. (PB-180-765)

WBTM T&EL 6 Final Report - The AMOS V Observer Aid Test, Part I. Elbert W. Atkins and Walter E. Hoehne, October 1968. (PB-180-547)

WBTM T&EL 7 Final Report - Test and Evaluation of the Fischer and Porter Precipitation Gage. Walter E. Hoehne, August 1968. (PB-180-290)

WBTM T&EL 8 Final Report - The AMOS V Observer Aid Test, Part II. Elbert W. Atkins and Walter E. Hoehne, March 1969. (PB-183-810)

WBTM T&EL 9 Analysis of Visibility Observation Methods. Frederick C. Hochreiter, October 1969. (PB-188-327)

NOAA Technical Memoranda

NWS T&EL 10 Analysis of Cloud Sensors: A Manual Height Measurement System. Staff, Observation Techniques Development and Test Branch, March 1971. (COM-71-00549)

NWS T&EL 11 Discussion of Sensor Equivalent Visibility. Staff, Observation Techniques Development and Test Branch, July 1971. (COM-71-00964)

NWS T&EL 12 Standardized Functional Tests. Walter E. Hoehne, December 1971.

NWS T&EL 13 Evaluation of Common Ceilometer Technology. Staff, Observation Techniques Development and Test Branch, December 1971.

NWS T&EL 14 (REREX) Remote Readout Experiment for Clouds and Visibility. Staff, Observation Techniques Development and Test Branch, December 1973. (COM-74-10535)

NOAA Technical Memorandum NWS T&EL-15

PROGRESS AND RESULTS OF FUNCTIONAL TESTING (SUPPLEMENT TO NOAA TECHNICAL MEMORANDUM NWS T&EL-12)

Walter E. Hoehne

Test and Evaluation Laboratory
Sterling, Va.
April 1977

UNITED STATES
DEPARTMENT OF COMMERCE
Juanita M. Kreps, Secretary

NATIONAL OCEANIC AND
ATMOSPHERIC ADMINISTRATION
Robert M. White, Administrator

National Weather
Service
George P. Cressman, Director

CONTENTS

PROGRESS AND RESULTS OF FUNCTIONAL TESTING
SUPPLEMENT TO NOAA TECHNICAL MEMORANDUM NWS T&EL-12

Walter E. Hoehne
National Weather Service, Test and Evaluation Division,
NOAA, Sterling, Va.

ABSTRACT. The procedures developed to perform a standardized functional test are described in detail with a step-by-step example. A table of precision and comparability of meteorological measurements determined thus far is included.

1. BACKGROUND

The adequacy of proposed new equipment is determined for the National Weather Service (NWS) by the Office of Technical Services (OTS) Test and Evaluation Division (T&ED). In addition to questions of accuracy and general utility, one question that must be answered is: "What change will there be in data provided to the user when a new system is adopted?" To answer this question, comparison is made between the output of the new system and the output of systems already in use. A program to standardize the evaluation of such differences in output from meteorological instruments was described in NOAA Technical Memorandum NWS T&EL-12. This report is a supplement describing the exact procedures followed in performing these standardized functional tests and giving a list of the results obtained to date.

Functional precision is the root mean square (rms) of the difference between readings from two (or more) identical sensors operating in the same environment. Usually the sensors are operated side-by-side in the outdoor environment. Readings are taken simultaneously from both sensors over a period of time, usually several months. The difference between corresponding readings is calculated and the rms of these differences is called functional precision.

The operational importance of functional precision is that it tells you when different readings from different stations using the same type sensor are significant. If two stations report wind directions differing by 10° and the functional precision of the instruments has been determined to be 15°, we can assume that the difference may not be a real difference in wind direction but rather is simply the result of the precision of the instruments.

Comparability is calculated the same way as functional precision. That is, they're both the rms of the differences in readings from the items being compared. Comparability is used to determine if there will be a significant change in the data when a new sensor is introduced into the NWS operating system.

2. STATISTICAL PARAMETERS

In T&EL-12 specific statistical parameters utilized in the standardized functional tests are the mean difference b, called the bias, which is used in evaluating the calibration and calibration procedures; the standard deviation about the mean difference $\sigma_{a1,a2}$ which is used as a measure of variability of output, and the functional precision $\zeta_{a1,a2}$ when a comparison is made between identical instrumentation, or comparability $\zeta_{A,B}$ when the comparison is made between different instruments designed for measuring the same atmospheric parameter.

$$b_{a1,a2} = \frac{\sum_{i=1}^{n} (X_{a1} - X_{a2})_i}{n}$$

where X_{a1} = Reading of one sensor

X_{a2} = Reading of an identical sensor

$$b_{A,B} = \frac{\sum_{i=1}^{n} (X_A - X_B)_i}{n}$$

where X_A = Reading of one sensor

X_B = Reading of a different sensor for measuring the same parameter

$$\sigma = \pm \left(\frac{\sum_{i=1}^{n} ((X - X) - b)_i^2}{n-1} \right)^{\frac{1}{2}}$$

$$\zeta_{a_1,a_2} = \pm \left(\sigma^2_{a_1,a_2} + b^2_{a_1,a_2} \right)^{\frac{1}{2}}$$

$$\zeta_{A,B} = \pm \left(\sigma^2_{A,B} + b^2_{A,B} \right)^{\frac{1}{2}}$$

We also calculate the third and fourth moments about the mean. These provide a measure of the skewness and peakedness of the distribution. Our evaluations and conclusions are based on the assumption of a normal distribution and the third and fourth moments provide a measure of the validity of that assumption. In general we don't report the values of the third and fourth moments, but refrain from publishing precision and comparability when the assumption of normal distribution doesn't appear to be valid.

3. PROCEDURES

The determination of the precision of a measurement was described in T&EL-12 as the root mean square (rms) of the difference observed between identical sensor systems exposed under "essentially the same conditions" and read out at the same time. "Essentially the same conditions" were defined as being within a horizontal distance less than or equal to 10 meters and a height difference as small as possible but in no case greater than 1 meter. The time difference between readings is minimized and is preferably less than one second, but because of the relatively slow time rate of change of most meteorological parameters, time differences of a minute or occasionally more may be tolerated.

3.1 Installation

For some measurements (e.g., visibility) the horizontal distance or the height (e.g., cloud height) may be the parameter of interest. In the first case, one of the two dimensions of horizontal distance is minimized and may not exceed 10 meters while all other criteria remain the same. In the second case all criteria for position and sampling remain unchanged and the measured height is treated as if it were an atmospheric parameter.

The physical dimension of some measuring systems may exceed the critical dimension indicated (e.g., a rotating beam ceilometer with an 800' baseline). In those cases the corresponding portions of the system (e.g., the detectors and projectors) are installed within the distance and height limits.

The bias as determined from the mean difference is not reported if it is less than one increment of resolution. If it is equal to or more than one increment of resolution it will be reported along with the reason for the bias if it can be established (e.g., the temperature reported by the lower of two radiosondes on the same balloon train is biased toward higher readings because of its location).

3.2 Sampling

When the systems (two or more) have been installed according to the previous instructions, samples are taken in pairs. The samples are taken simultaneously with a minimum time difference between the sampling of one instrument system and the sampling of another. The time interval between the pairs of samples must be long enough to insure that the samples are indeed independent. If, for example, a particular temperature sensor has a time constant of 5 minutes or more, taking pairs of samples at time intervals of less than 5 minutes would not produce an independent sampling procedure. The nature of atmospheric data is such that time intervals between pairs of samples as long as an hour or more may be desirable.

The number of samples taken, in general, is in excess of 100 and should be as large and cover as wide a range of atmospheric conditions as practical. The number of samples cannot be too large. To make certain that we have enough samples, we have established the criteria from the 99.7% confident that the difference Δ between the derived mean of a set of samples and the true mean of the population of all samples is less than or equal to three times the standard deviation σ about the mean, divided by the square root of the number of samples in our set:

$$\Delta < \frac{3\sigma}{n^{\frac{1}{2}}} .$$

We don't consider the sampling complete until Δ is less than or equal to one increment of resolution of the system being tested. Or stated another way, the number of samples needed is

$$n \geqq \frac{(3\sigma)^2}{\Delta} .$$

3.3 Correlation

Particular differences may change with the size of the measurement (e.g., the rms difference in the measurement of wind speed by two systems may be greater or smaller at high wind speeds than at low wind speeds) and the data is tested for such correlation by obtaining the best linear, power, or exponential fit via the least squares method. If the correlation coefficient is < 0.3 it is assumed that the correlation is not significant and it is not reported. If the correlation coefficient is $\geqq 0.3$ the functional relationship is reported.

Similar correlation between the differences and another parameter may be present. (e.g., rms difference in temperature measurement by radiosonde may be correlated with the pressure at which the measurement is made.) If such correlations are suspected or postulated, they are investigated and reported according to the same criteria as for correlation with size.

The correlation between the size of the difference and the measurement may be discontinuous. In such cases frequency distributions are used to evaluate the differences. Two wind direction sensors may show large differences when measuring one particular direction and relatively small differences at other times. A frequency distribution is well suited to the disclosure of such a feature but requires the accumulation of a large number of samples.

3.4 Example

Functional tests were made to establish the precision and comparability of 24-hour precipitation accumulation data obtained from the U.S. 8-inch precipitation gage and the WMO Snowden pit gage.

3.4.1 Installation

The gages were installed within 10 meters of each other in the same manner as would be used in the field.

3.4.2 Sampling

Readings were made at the same time each day for a period of 3 years and samples were compared in pairs for each measurement. The samples were labeled as follows:

R_1 = One U.S. 8-inch gage

R_2 = Other U.S. 8-inch gage

$A_{1,2}$ = Snowden pit gages

3.4.3 Statistical Analysis

The sums:

$$S_{1,j} = \sum_{i=1}^{n_1} (R_1 - R_2)_i^j, \quad S_{2,j} = \sum_{i=1}^{n_2} (A_{1,2} - R_1)_i^j,$$

and $$S_{3,j} = \sum_{i=1}^{n_3} (A_{1,2} - R_2)^j$$ were formed.

$n_{1,2,3}$ = the number of samples

j = 1, 2, 3, 4

When: j = 1 this used to produce the bias

j = 2 the square root produces the rms

j = 3 the skewness

and j = 4 the peakedness.

The calculations produced:

$$\text{the functional precision } \zeta_{a_1,a_2} = \left(\frac{S_{1,2}}{n}\right)^{\frac{1}{2}} = \pm 0.3\text{mm}$$

$$\text{the mean difference } b_{a_1,a_2} = \frac{S_{1,1}}{n} = \pm 0.03\text{mm}$$

$$\text{the comparability } \zeta_{A,B} = \left(\frac{S_{2,2} + S_{3,2}}{n_2 + n_3} \right)^{\frac{1}{2}} = \pm 1.1\text{mm}$$

$$\text{and the bias } b_{A,B} = \frac{S_{2,1} + S_{3,1}}{n_2 + n_3} = +0.4\text{mm.}$$

The third and fourth moments were calculated and found to be 0.26 and 10.16 in the first case and -7.14 and 81.27 in the second. In the first case the distribution is bell shaped but peaked. The precision as reported is therefore conservative. In the second case the distribution is peaked and skewed. The skewness is a result of the correlation between the difference observed and the size of the measurement. Thus the comparability is reported as a function of the measurement.

The standard deviation about the mean difference is:

$$\sigma_{a_1,a_2} = \pm \left(\zeta^2_{a_1,a_2} - b^2_{a_1,a_2}\right)^{\frac{1}{2}} = \pm 0.3\text{mm.}$$

From this we find:

$$n \geqq \frac{(3\sigma)^2}{\Delta} = 8$$

will provide enough samples for confidence in the mean. ($\Delta = 0.1$mm which is one increment of resolution.) For this measurement we had 149 samples. A similar calculation shows that 109 samples would be needed for confidence in the mean of the comparability (i.e., the bias of the pit gage when compared to the 8-inch gage). We had 594 samples.

Using the difference between samples as one argument and the size of the sample from one of the gages as the other, a computer program was used to calculate the best least-squares linear, power, and exponential curve fit to the precision data. The best fit was the linear relation and the correlation coefficient was $0.24 \leq 0.3$ and therefore not reported.

The mean difference was -0.03 which is much smaller than one increment of resolution and therefore not reported.

The bias of the pit gage as compared to the 8-inch gage is 0.4mm with the pit gage reading higher. A strong correlation was found between the size of the difference and the total 24-hour accumulation. The linear correlation coefficient was 0.72. The linear equation fitted to the data was $d = 0.08T \pm 0.4$mm where d is the difference expected and T is the total 24-hour accumulation.

4. SUMMARY

A. The systems to be compared are installed within a volume not more than 10 meters in horizontal distance and 1 meter in vertical extent.

B. Measurements are made simultaneously.

C. The measurements are compared in pairs with a time interval between pairs of measurements at least twice the time constant of a particular measuring instrument.

D. The root mean square of the differences is calculated to provide functional precision if the systems are the same and comparability if the systems are different.

E. The maximum number of pairs of measurements practical will be obtained with no less than N pairs utilized for a precision determination where $N \geq \frac{3\sigma^2}{\Delta}$, σ = standard deviation, Δ = one increment of resolution.

F. Data is tested for correlation with size and other parameters. Functional relationship is reported if correlation is ≥ 0.3.

G. Bias or mean difference is reported if it equals or exceeds one increment of resolution. The reason for the bias is reported if it can be determined.

APPENDIX A. PRECISION AND COMPARABILITY OF METEOROLOGICAL DATA

Precision and Comparability of Meteorological Data

Surface Sensors

Parameter Measured	Sensor	Functional Precision
Temperature	HO-61	$\pm 0.8^{\circ}$F
Dew Point	HO-61	$\pm 0.9^{\circ}$F
Wind Speed	F 420	± 1.0 kt
Wind Speed	DARDC	± 0.8 kt
Wind Direction	F 420	$\pm$ 5°
Wind Direction	DARDC	$\pm 22.5^{\circ}$
24-Hour Precipitation	Weighing rain gage	$\pm$0.02"
24-Hour Precipitation	U.S. 8-inch gage	$\pm$0.3mm
24-Hour Precipitation	Snowden pit gage	$\pm$0.7mm
24-Hour Maximum Temperature	Mercury-in-glass thermometer Cotton Region Shelter	$\pm 0.2^{\circ}$F
24-Hour Maximum Temperature	Mercury-in-glass thermometer Israeli Shelter	$\pm 0.2^{\circ}$F
24-Hour Minimum Temperature	Mercury-in-glass thermometer Cotton Region Shelter	$\pm 0.2^{\circ}$F
24-Hour Minimum Temperature	Mercury-in-glass thermometer Israeli Shelter	$\pm 0.2^{\circ}$F
Altimeter Setting (Pressure)	Kollsman Aneroid Barometer	$\pm$0.000" Hg.
Peak Wind Speed	DARDC	$\pm$ 4 kt
Solar Radiation	Eppley Pyranometer + Monitor Labs Integrators	$\pm$0.35 langley/hr 4.1 Wm^2
Visibility	Videograph	$\pm$0.2 mi

Upper Air Sensors

Parameter Measured	Sensor	Functional Precision	Bias
Temperature	WBRT + J005 Sonde	$\pm 0.5^{\circ}C$	$0.2^{\circ}C$*
Dew-Point Depression	"	$\pm 3.6^{\circ}C$	$-0.9^{\circ}C$*
Pressure	"	± 1.9 mb	
Wind Speed	"	± 3.1 mps	
Wind Direction	"	$\pm 14.5^{\circ}$	
Wind Vector	"	2.8 mps	

*Lower sonde on balloon train affected by heat and moisture from water-activated battery of upper sonde.

		Functional Precision
Cloud Height	Rotating Beam Ceilometer	$\pm 0.01309 \dfrac{B^2 + h^2}{B + 0.01309h}$

B = Baseline
h = Measured Height

		Comparability	Bias
Temperature	HO-61/AMOS III-70	$\pm 0.7^{\circ}F$	
Dew Point	HO-61/AMOS III-70	$\pm 0.8^{\circ}F$	
Wind Speed	F 420/AMOS III-70	± 1.0 kt	
Wind Direction	F 420/AMOS III-70	$\pm 7^{\circ}$	
Wind Direction	DARDC/AMOS III-70	$\pm 18^{\circ}$	11°
24-Hour Maximum Temperature	Israeli Shelter vs. Cotton Region Shelter	$\pm 0.4^{\circ}F$	$-0.2^{\circ}F$
24-Hour Minimum Temperature	Israeli Shelter vs. Cotton Region Shelter	$\pm 0.4^{\circ}F$	$-0.2^{\circ}F$
24-Hour Precipitation	U.S. 8-inch gage vs. Snowden pit gage	$0.08T \pm 0.4$mm T = 24-Hour Precipitation	