NOAA Technical Memorandum NWS T&EL-12

U.S. DEPARTMENT OF COMMERCE
National Oceanic and Atmospheric Administration
National Weather Service

Standardized Functional Tests

WALTER E. HOEHNE

Systems
Development
Office

Test and
Evaluation
Laboratory

STERLING, VA.

December 1971

NOAA Ti

National Weather Service, oratory Series

The Test and Evaluation Laborato fice of Systems Development conducts experiments and tests to dete ity of newly developed equipment, instruments, and systems for field in by the National Weather Service (NWS). Results of tests and evaluation rovided to the other operating Offices of the NWS (Office of Meteorological Ope ions, 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 to 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.

NOTE--The National Weather Service does not approve, recommend or endorse any proprietary product or proprietary material mentioned in this publication. No reference shall be made to the National Weather Service, or to this publication furnished by the National Weather Service, in any advertising or sales promotion which would indicate or imply that the National Weather Service approves, recommends or endorses any proprietary product or proprietary material mentioned herein, or which has as its purpose an intent to cause directly or indirectly the advertised product to be used or purchased because of this National Weather Service publication.

Publications listed below are available from the National Technical Information Service, U.S. Department of Commerce, Sills Bldg., 5285 Port Royal Road, Springfield, Va. 22151. Price: $3.00 paper copy; $0.95 microfiche. Order by accession number shown in parentheses at end of each entry.

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)

(Continued inside back cover)

U.S. DEPARTMENT OF COMMERCE
National Oceanic and Atmospheric Administration
National Weather Service

NOAA Technical Memorandum NWS T&EL-12

STANDARDIZED FUNCTIONAL TESTS

Walter E. Hoehne

Systems Development Office
Test and Evaluation Laboratory

Sterling, Va.
December, 1971

UDC 551.508.001.4:389.6

551.5	Meteorology
.508	Meteorological instruments
.001.4	Evaluation and testing
389.6	Standardization

TABLE OF CONTENTS

STANDARDIZING FUNCTIONAL TESTS

Walter E. Hoehne

ABSTRACT

A standardized test for evaluation of meteorological measuring systems in the natural environment is described. The test is designed to provide a quantitative statistical value that will indicate the reliability of a particular system output. This quantity called functional precision provides a quantitative estimate of the difference in readings that can be expected from systems of identical design and construction when exposed to the same environmental conditions. The mathematical definition of this parameter is described, methods of application are discussed, and a specific example is presented.

Background

The adequacy of proposed new equipment is determined for the National Weather Service by the Systems Development Office, Test and Evaluation Laboratory. 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 a system already in use. The Functional Experimentation and Test Branch has developed a program to standardize the evaluation of differences in output from meteorological instruments. Functional is used here to indicate tests made with the equipment being operated in the natural environment and not under controlled laboratory conditions.

The value of a particular measurement for meteorological purposes has, in most cases, not been objectively defined. Some efforts are being made to make such definitions. For example, the WMO Commission for Instruments and Methods of Observation (CIMO) has appointed a committee to define temperature for meteorological purposes. Physical measurements may be defined in terms of physical phenomena (e.g., the phase changes of water were chosen as two points on a temperature scale). Meteorological measurements are not so clearly defined. A measurement for meteorological purposes is associated with physical conditions, with a volume larger than the volume immediately in contact with the sensor and an arbitrary time period. Perlat and Petit[1], and Bragenskaia and Kagan[2,3] have investigated the problems of associating instantaneous point measurement with a time and/or space domain. Many investigators have addressed themselves to the accuracy of particular instruments and laboratory methods for determining accuracy. Recently Beckman[4] proposed a means of setting up standards for instrument parameter definitions and for test procedures. Lamb and Pharo[5] also proposed standardizing meteorological instrument testing. In both cases, the proposed tests are those to be conducted in a laboratory with a controlled environment that simulates the natural environment. The variation in reading due to the natural variability of the atmosphere can be considered only to the extent that such variability can be simulated.

The Meteorological Working Group (MWG) Inter-Range Instrumentation Group (IRIG) compiled a set of accuracies for meteorological equipment used on the National Missile Ranges[6]. The MWG has revised that document[7] and in it a new concept is employed expressing "reliability" of data rather than accuracy. Quote; "Reliability is defined as the best available quantitative estimate of the quality of the data for operational use at the test ranges. Where possible, and as noted, the term reliability includes errors resulting from human and instrumental sources. Where standards have been established, reliability is a statement of accuracy. In general, however, the values of reliability are statements of data precision to be expected from well maintained equipment, operated by competent individuals according to a well defined procedure." The program described here is an attempt to standardize one test for reliability of meteorological data.

Objective

In the past when a new sensor system was developed, its data reliability was evaluated by comparing it with an existing system. Output differences between the two systems were tabulated and analyzed statistically to produce mean difference, variance, et cetera. These statistical results could not be evaluated because no information was available on the difference that could be expected due to natural variability. It was not known if the difference between the two systems was the same as, larger than, or smaller than that which would have been observed had the systems been alike. We developed a concept called functional precision to provide a measure of the difference that could be expected. We expanded this concept to include a measure of accuracy.

The definitions contained in the MWG Document[7] are adhered to. Of special interest are the definitions of accuracy and precision. According to Websters Dictionary these words are synonymous, but the IRIG Glossary of Terms[8] makes a distinction as follows:

ACCURACY

The numerical difference between any value and the true value. Applied by transference to the instrument or system producing the value. Distinguished from PRECISION.

1. Instrumental - The accuracy of a measurement after the errors caused by elements external to the instrument are removed. A measure of the accuracy of the instrument proper.

2. Transducer - The ratio of the error to the full-scale output (expressed as "within ±----- percent of full-scale output") or the ratio of the error to the output, expressed in percent.

PRECISION

1. A measure of the reproducibility with which one instrument (or several instruments of the same type) can reproduce repeatedly measurements of the same quantity. If the precision is high, such results will lie within a narrow range.

2. Adapted for extremely accurate scientific measurements. It is not, however, a guarantee of accuracy (negligible error), because precision refers to the measuring instrument and does not cover external sources of error inherent in the measuring method.

3. Computation - The degree of exactness with which a quantity is stated, as contrasted with ACCURACY, which is the degree of exactness with which a quantity is known or observed.

4. Of Measured Data or of a Measuring Instrument - In general, the uniformity of data from repeated measurements of the same constant phenomenon. In the case of a constantly changing phenomenon, the word precision has a similar meaning. The best measure of precision in the latter case is the STANDARD ERROR OF ESTIMATE (S). The smaller S is, the higher the precision. Precision usually is a function of the time interval between measurements and so should be qualified. See ACCURACY.

Program expansion required the designation of standards, but there is very little agreement on the relative merits of various standards. To avoid being involved in extensive defense of standards chosen, we accepted standards designated by others with the requirement that the designator provide documentation of laboratory accuracy, instructions on calibration and maintenance and duplicate systems. The precision of the standard is measured by the test described in this report and other measuring systems are compared with the standard either concurrently or subsequently. In every case duplicate systems are tested to determine precision as defined[7,8]. According to this definition a measuring system can be precise without being accurate, but not conversely.

Statistical Concepts

The statistical definition and algebra for manipulation are based on the work of Ku[9]. He defines σ as an index of precision where σ (the standard deviation) is the positive square root of the variance

$$\sigma_x^2 = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \left(X_i - \overline{X_i}\right)^2$$

The X_i's are repeated measurements by a system of a single physical quantity under essentially the same conditions.

Meteorological quantities in the uncontrolled natural environment do not permit such repetitive measurements. They change with time, hence sequential measurements would be contaminated by diurnal and climatic changes and not be useful for system evaluation.

If X_i and X_j are independent readings of the same physical quantity under essentially the same conditions, then

$$(1) \quad \sigma_{xi+xj}^2 = \sigma_{xi}^2 + \sigma_{xj}^2 = \sigma_{xi-xj}^2$$

$$= \lim_{N\to\infty} \frac{1}{N} \sum_{i,j=1}^{N} \left(\left(X_i - X_j\right) - \left(\overline{X_i - X_j}\right)\right)^2$$

The natural environment does not permit sequential sampling by the same system, hence we use duplicate systems a_1 and a_2 to make independent simultaneous measurements of the same quantity under essentially the same conditions. The time and space sampling procedures described later define what we mean by essentially the same conditions.

Equation (1) becomes

$$(2) \quad \sigma_{a_1,a_2}^2 = \sigma_{xa_1}^2 + \sigma_{xa_2}^2$$

$$= \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \left(\left(X_{a_1 i} - X_{a_2 i}\right) - \left(\overline{X_{a_1} - X_{a_2}}\right)\right)^2$$

Although duplicate systems are used $\overline{X_{a_1}}$ may not equal $\overline{X_{a_2}}$ that is, there may be a bias between duplicate systems and this must be accounted for. The size of the bias b is determined from the mean of the differences.

$$(3) \quad b = \left(\overline{X_{a1} - X_{a2}}\right) = \frac{1}{N}\sum_{i=1}^{N}(X_{a1i} - X_{a2i})$$

Where: X_{a1} = the reading of one system

X_{a2} = the reading of a duplicate system at the same time and exposed within a predetermined distance from the first system

Substituting from (2) and (3) and expanding

$$\sigma_{a1,a2}^{2} = \frac{1}{N}\sum_{i=1}^{N}\left((X_{a1i} - X_{a2i})^{2} - 2(X_{a1i} - X_{a2i})b + b^{2}\right)$$

$$= \frac{1}{N}\sum_{i=1}^{N}(X_{a1i} - X_{a2i})^{2} - 2b\left(\frac{1}{N}\sum_{i=1}^{N}(X_{a1i} - X_{a2i})\right) + b^{2}$$

$$(4) \quad \sigma_{a1,a2}^{2} = \frac{1}{N}\sum_{i=1}^{N}\left(X_{a1i} - X_{a2i}\right)^{2} - b^{2}$$

Following the recommendation of ASTM[10] we modified the word "precision" with the word "functional" to indicate determinations made under operational conditions in the natural environment. We define "functional precision" of system design a as:

$$(5) \quad 6_{a1,a2} = \pm\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(X_{a1i} - X_{a2i}\right)^{2}}$$

Substituting this in equation (4) and rearranging terms produces

$$(6) \quad 6_{a1,a2} = \pm\sqrt{\sigma_{a1,a2}^{2} + b^{2}}$$

$6_{a1,a2}$ includes the instrumental and environmental random variability $\sigma_{a1,a2}$ and the systematic bias b. The data are used to obtain $6_{a1,a2}$ from equation (5), b from equation (3) and $\sigma_{a1,a2}$ is obtained from

$$(7) \quad \sigma_{a1,a2} = \pm \sqrt{6_{a1,a2}{}^{2} - b^{2}}$$

Time and Space Sampling

To define what is meant by "essentially the same conditions" we note first that meteorological parameters may change with time. The time Δt between X_{a1} and X_{a2} should, therefore, be as small as possible. The size of the time interval is governed by the sophistication of our data gathering methods, but is in most cases less than one second.

The spatial separation Δl between the readings X_{a1} and X_{a2} may be resolved into three components. Only two of these can be reduced to zero. Because present meteorological practice reports synoptic measurements referenced to horizontal surfaces, the vertical component of Δl is reduced theoretically to zero. The actual vertical component depends on the ability to measure and control the vertical position of the sensors. No attempt is made to distinguish between the two horizontal components. Instead, the magnitude of the resultant is limited by specifying a horizontal plane circle with a maximum diameter $\leq$ 10 meters. Again, the separation is reduced to the practical minimum permitted by ability to measure and control the distance and the necessity of avoiding interaction between sensors.

A slight change in methodology is required to determine the functional precision of upper air systems. To obtain comparative readings from instruments that are carried aloft by a balloon, two radiosondes are attached to the train. A distance of five meters is used as the vertical separation. The reading from the upper radiosonde at some time T is compared with reading of the lower radiosonde at a time T + 5/A where A is the ascent rate of the balloon train. At presently used ascent rates (approximately 300 meters per minute) $5/A \approx 1$ second. The horizontal separation between the measurement made by the lower unit and the measurement made by the upper unit will be more than 10 meters in many instances. The balloon train travels with the wind, however, so that the separation with respect to the horizontal air stream will be less than 10 meters. The bias observed may be correlated with the relative position of the radiosondes in the balloon train. The data are analyzed to detect such correlation and the bias resulting from position is subtracted.

The definition of functional precision is based on a large number of samples (i.e., $N \to \infty$ equation (1)). When N is large so that $\frac{1}{N-1} \approx \frac{1}{N}$ we use equations (6) and (7). For smaller samples, equations using $\frac{1}{N-1}$ instead of $\frac{1}{N}$ should be used. Measurement should be made of as many values of each parameter as possible, and data gathering period for each determination may extend over several seasons so that in most cases N>>100. Automatic data logging equipment reads and records the output from sensors exposed during a particular data gathering program. Using this equipment, the sample N can be made very large without large expenditures of manpower. The automatic data logger samples the sensors simultaneously and records the data in digital form suitable for computer processing. Sampling is repeated as often as possible to ensure a large value of N, but is not repeated with a frequency that destroys the independence of the individual sample.

Application

The most important factor in an evaluation of output data is accuracy. To measure accuracy, a knowledge of the true value is required. Such knowledge is usually not available for meteorological measurements. We use the statistical methods described in this report as a substitute for comparison with a known true value. Using equations (3), (5) and (7) a comparison between standard A and system B produces comparability 6,A,B, bias $(\overline{X_A - X_B})$ and $\sigma A,B$ which can be evaluated by comparing them with the functional precision $6A_1,A_2$, bias $(\overline{X_{A1} - X_{A2}})$ and $\sigma A_1,A_2$ of the standard. Similar comparisons are made between standards with the same laboratory accuracy, but designated by different sources. If $6B_1,B_2 < 6A_1,A_2$ the data from standard B will be operationally more reliable than the data from standard A. This assumes that data are obtained only when both systems are calibrated and properly maintained.

Application Problems

Few problems occur when the methods described above are applied to instruments that make an instantaneous point measurement of parameters that are continuous in time and space. The temperature measurement, for example, is continuous throughout the time and space domain to which it is applied. The fact that temperature measurements are statistical expressions of the kinetic energy of molecules does not pose a problem because the time and space domains involved are much smaller than those of interest for meteorological purposes. Application to measurements that involve integration over longer times is more difficult. For example, the measurement of total precipitation is the integration of precipitation rate over some time period. The integration is done in different ways depending upon the particular instrument involved. It may be done by summing incremental amounts or by a single direct measurement of total accumulation. The precipitation may be zero in some portions of the time and space domain of interest, but have large values in other portions, with sharp discontinuities between the two areas. High sampling rates are not applicable because the integration time is the same as the time interval of interest. An extended program will probably be required to obtain the number of samples needed to determine the functional precision.

Occasionally the functional precision will be correlated with the numerical magnitude of the readings. For example, either the random difference, the bias or both may be larger when two anemometers are reading high wind speeds than when the same two are reading low wind speeds. The data from each test is examined to detect such correlation. Such correlation will not be reported when the correlation coefficient is $\leq .3$ indicating that less than 10% of the variation in the difference between readings can be explained by a relationship between difference and magnitude of the readings. The equations relating the random difference and/or the bias to the magnitude will be provided when the correlation coefficient exceeds .3. The functional precision and bias will be presented in graphical form when observed differences vary systematically as the magnitude of the observed value. This graphical presentation is most important when the bias is small or negative in one range of observed values and large or positive in another.

Example of Use

When new observing equipment is proposed which is intended to produce an output different from, but related to, current operational practice the user should be informed as to how data from the new system will differ from that provided by the existing system. A recent test was made of a system that reports automatically the peak one second wind speed during the past hour. This speed is to be used for the same purposes as the speed reported for maximum gust is presently used. The latter is obtained by manual reading of a wind speed recorder record. The records were read by trained observers and the comparisons described earlier were made. The functional precision of current operational practice $\mathscr{6}A_1,A_2 = \pm.402$ knots. The bias $b_{A_1,A_2} = .04$ knots and $\sigma A_1,A_2 = \pm.4$ knots. The comparability between observers reading the recorder and the automatic output $\mathscr{6}A,B = \pm.67$ knots, the bias $b_{A,B} = -.3$ knots and $\sigma_{A,B} = \pm.6$.

The comparability can be evaluated by noting that it is of the same order of magnitude as the functional precision of current practice and both are smaller than the resolution (1 knot) and the required accuracy (±2 knots) of the systems. Note also that $2\mathscr{6}A,B < 2$ knots. If current practice is designated as the standard and assumed to be "true" then the data from the automatic system is within the accuracy limits required. The small bias observed between current practice and the automatic system can be attributed to slight differences in the structure of the sensor. This small bias is not significant when compared to the resolution and required accuracy of the systems. Users have been informed that there will be no significant change in the data when the automatic system replaces current practice.

Program Plan

The determination of functional precision and bias can continue as long as new sensor systems are developed. At present, we are making an effort to determine the values for the sensor systems most widely used by the National Weather Service. The time required for each determination will depend upon the exact nature of the sensor and the character of the measurement. As pointed out above, determination for sensors making certain types of measurements (i.e., precipitation, evaporation, visibility, cloud cover, solar radiation) will take long periods of low level effort while determination for other types (i.e., temperature, dewpoint, wind speed, wind direction) can be accomplished in relatively short periods of concentrated effort.

Determinations for some surface meteorological systems is now under way. Specifically these include instrument shelters, hygrothermometers and rain gages. We combine these determinations with other test and evaluation programs to make full use of the data. Anemometers will be added to the surface instrumentation evaluation in the near future.

We have started the determination of functional precision and bias for the presently used upper air system. This determination will require approximately one year. We will determine the functional precision and bias of the next generation upper air system as part of the test and evaluation of that system.

The final report on the system will include the determination for both the new and the old systems. The users of the data will then have quantitative information on how the new data differs from what he received previously.

REFERENCES

1. Perlat, A.,and Petit, M., Mesures en Meteorologie, (Measurements in Meteorology.) Editor, A. Viaut, Monographies des Meteorologie, Gauthier-Villars, Paris, 1961, 393 pp.

2. Braginskaia, L.L., and Kagan, R.L., "O tochnosti opredeleniia srednikh velichin po mgnovennym, otschetam inertsionnykh priborov," (Accuracy in Determining Mean Values from Instantaneous Readings of Inertial Instruments.) Trudy, No. 14, Glavnaia Geofizichesk ara Observatoriia, Leningrad, 1965, pp 3-20.

3. Braginskaia, L.L., and Kagan, R.L., "O vliianii inertsii prebara na tochnost' meteorologicheskikh izmerenii, (Effect of Instrument Inertia on the Accuracy of Meteorological Measurements.) Trudy, No. 191, Glavnaia Geofizicheskaia Observatoriia, Leningrad, 1966. pp. 72-76.

4. Beckman, John C., "On the Need for Standardization of Instruments Calibration and Related Terminology," Bulletin of the American Meteorological Society, Vol. 48, No. 9, Sept. 1967, pp. 704-705.

5. Lamb, Robert C., and Pharo, James A., "Standardizing Meteorological Instrument Testing," Bulletin of the American Meteorological Society, Vol. 48, No. 9, Sept. 1967, pp. 707-708.

6. Meteorological Working Group IRIG, RCC, "Meteorological Equipment Data Accuracies" Document 110-64, Secretariat, Range Commanders Conference, White Sands Missile Range, N. Mex., Feb. 1965, 11 pp.

7. Meteorological Working Group, IRIG, RCC, "Reliability of Meteorological Data" Document 110-70, Cancelled, Secretariat, Range Commanders Conference, White Sands Missile Range, N. Mex., 1970.

8. U.S. Inter-Range Instrumentation Group Range Commanders Conference, Glossary, (Draft), Secretariat, Range Commanders Conference, White Sands Missile Range, N. Mex., Sept. 1963, 494 pp.

9. Ku, Harry H., "Statistical Concepts in Metrology," The Handbook of Industrial Metrology, American Society of Tool and Manufacturing Engineers, Prentice Hall, Inc., New York 1967, pp 20-50.

10. "Use of the Terms Precision and Accuracy as Applied to the Measurement of a Material", Book of ASTS STANDARDS, E177-61T, American Society for Testing and Materials, Philadelphia, Pa., 1961, Various pp.

(From inside front cover)

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