**7.12** IMPROVED REAL-TIME QUALITY CONTROL OF NDBC MEASUREMENTS

David B. Gihousen*
National Data Buoy Center, Stennis Space Center, MS 39529-6000

## 1. INTRODUCTION

Data quality control is not the most interesting or popular topic at an American Meteorological Society conference. That is probably why the session is always relegated to a Friday morning! Nevertheless, most operational meteorologists can recall times when an embarrassingly bad observation was transmitted from an automatic weather station to the public. More serious are erroneous, but apparently reasonable observations that might lead to an incorrect operational decision. So, in that context, preventing these "mishaps" becomes very important.

The National Data Buoy Center (NDBC) has an extensive program to reduce greatly the chances of transmitting degraded measurements. These measurements are taken from its network of buoys and Coastal-Marine Automated Network (C-MAN) stations (Meindl and Hamilton, 1992). This paper discusses improvements to the real-time software that automatically validates the observations. In effect, these improvements codify rules data analysts—who have years of experience with our system—use to detect degraded data.

## 2. CONTEXT OF REAL-TIME QUALITY CONTROL

Before delving into the details, the proper place of these algorithms in the data quality control process needs to be understood. They form the last line of defense in preventing degraded data from reaching the public. Many other important measures precede them. New sensors are tested in environmental chambers before field evaluations. When the onboard software is modified, it is regression tested to make sure the output agrees with previously accepted values. Measurements from new sensors, buoy hulls, or onboard systems, called payloads, are compared with measurements from standard configurations (Giohousen, 1987). All sensors are calibrated before every deployment, and site surveys are conducted to ensure proper exposure of the anemometers at new C-MAN stations. Servicing technicians remain at the station until several hours of acceptable transmissions make it through the satellite.

The historical role of the real-time data validation algorithms was in removing the large, "gross" errors (Gilhousen, 1988). These errors are typically caused by such things as satellite transmission problems, power system degradation, and broken cable connections. What these algorithms detect is virtually certain to be wrong. Although these checks have done a poorer job of detecting errors due to sensor degradation, the modified algorithms are expected to perform much better. Furthermore, NDBC supplements the real-time checks with an extensive, after-the-fact validation effort at NDBC.

This effort at NDBC is a "manual-machine" mix that involves a different set of algorithms and a review of computer graphics. It is typically accomplished within 24 hours of observation time. When degraded data are detected, the analysts update a control file that instructs the real-time processing not to encode measurements from the offending sensor. NDBC is introducing many improvements to the process, which will be documented in a future paper.

## 3. EXISTING DATA VALIDATION TECHNIQUES

The validation methods used in 1996 will be presented before the modifications are described. The validation occurs via software running at the National Weather Service (NWS) Telecommunication Gateway (NWSTG) that encodes the observations into World Meteorological Organization (WMO) or NWS-approved codes. Measurements of sea level pressure, air temperature, sea surface temperature, dew point temperature, wind speed, wind direction, wind gust, wave height, average wave period, and dominant wave period are validated, if measured. If any measurement fails these checks, it will not be released, and measurements from a backup sensor, if it exists, will be examined. All NDBC stations have two anemometers; all buoys have two barometers; and a few buoys have two air temperature sensors.

Several transmission checks are accomplished before the data are validated. Any message with a single parity error is not transmitted. The wave portion of the message is transmitted in binary at the end of the transmission. If this message is shorter than expected, contains checksum errors, or has an improper synch character, no waves are encoded.

The simplest of the data checks, the range check, ensures all measurements fall within established upper and lower limits. A different set of limits is used in each of 29 climatologically similar areas.

The second check is the time-continuity check. The formula for performing the time-continuity check is:

$$M = 0.58\ \sigma\sqrt{\tau} \quad , \tag{1}$$

where $M$ is the maximum allowable difference, $\sigma$ is the standard deviation of each measurement, and $\tau$ is the time difference in hours since the last acceptable observation. $\tau$ is never greater than 3 hours despite the actual time difference. For information on how this formula was derived, see National Data Buoy Center (1996).

In practice, using station-specific values of the standard deviation of measured variables is not necessary. The general values in use are listed in Table 1. As with the general range limits, departing from the general values of $\sigma$ is necessary for some stations. For

---

* *Corresponding Author Address*: David B. Gilhousen, Data Systems Division, Building 1100, SSC, MS 39520-6000; e-mail: dgilhousen@ndbc.noaa.gov.

example, since water temperatures near stations close to the Gulf Stream can change abruptly, $\sigma^{WTMP}$ for several east coast stations was increased to 12.1 °C.

Four exemptions to the time-continuity test exist. These exemptions are based on the very rapid changes that occur in wind, pressure, temperature, and wave height during the passage of tropical cyclones and severe extratropical cyclones. First, air pressure measurements that fail the time-continuity check are released if both the previous and current pressures are less than 1000 hPa. Second, wind speed measurements are released if both the previous and current pressures are less than 995 hPa. Third, air temperature measurements are released if either the wind speed exceeds 7 m/s or the wind direction change is greater than 40°. Finally, wave height measurements are released if the current wind speed is equal to or greater than 15 m/s. Even with these contingencies in place, analysts can elect to disable the range- and time-continuity checks for limited periods during hurricanes.

**TABLE 1.** Standard deviations used for the time-continuity check

| Measurement | σ |
|---|---|
| Sea Level Pressure | 21.0 hPa |
| Air Temperature | 11.0 °C |
| Water Temperature | 8.6 °C |
| Wind Speed | 25.0 m/s |
| Wave Height | 6.0 m |
| Dominant Wave Period | 31.0 s |
| Average Wave Period | 31.0 s |
| Relative Humidity | 20.0% |

Finally, internal-consistency checks are:

- If the battery voltage is less than 10.5 volts, pressure is not released. This precludes the chance of transmitting bad pressures from a failing station operating on minimum power.
- The significant wave height, average, and dominant wave periods are set to zero if the significant wave height is less than 0.15 m. Without this, unrepresentatively large wave periods could be transmitted from an essentially flat, “signal-less” spectrum.
- If the dew point exceeds the air temperature by less than 1.1 °C, the dew point is set equal to the air temperature. If the dew point exceeds the air temperature by more than 1.1 °C, the dew point is not encoded. This approach is taken because a reading slightly more than 100 percent (Breaker et. al, 1997) is normal for the hygrometers used by NDBC.
- If the ratio of gust-to-mean wind speed is greater than 4 or less than 1, neither the wind speed nor gust is transmitted.

## 4. VALIDATION CHANGES FOR BETTER QUALITY CONTROL

Several changes will be made in early 1998 that will help further reduce the chance of degraded data being transmitted:

- If a measurement fails either the range or time-continuity check for two consecutive observations, the measurement is not transmitted until it is manually reviewed at NDBC.
- A check was installed to ensure that consecutive 10-minute wind direction averages, on stations equipped with “continuous winds,” agree with the standard 8-minute average wind direction. More specifically, the 10-minute average wind direction that overlaps the standard one must agree within 25° of the standard if the wind speeds exceed 2.5 m/s.
- A procedure was installed to determine if measurements from duplicate sensors in reasonable agreement. If the measurements fail this determination, the software transmits the measurement from the sensor that exhibits better time continuity. This sensor is then chosen for all subsequent reports until it is manually reviewed. If the measurements from duplicate sensors are within the tolerances listed in Table 2, they are judged to be in reasonable agreement. This procedure is designed to detect significant sensor failures automatically and switch to a backup sensor, if one exists and is transmitting reasonable values.

Figure 1 is a time-series plot of a wind speed failure. Under the older procedure, almost 24 hours of degraded data were transmitted before data analysts at NDBC detected a wind speed failure and manually switched to the other sensor. With this new procedure, the software would automatically transmit speeds from the second sensor beginning at 1500 UTC, on March 1, 1996.

TABLE 2. Tolerances used to validate duplicate measurements

| Measurement | Tolerance |
|---|---|
| Wind Speed | 1.5 m/s |
| Wind Direction | 25° provided speed > 2.5 m/s |
| Sea Level Pressure | 1.0 hPa |

The number of degraded measurements transmitted are estimated to fall from approximately 40 to 15 each month. Excluding spectral wave measurements, about 800,000 measurements are transmitted each month.

## 5. VALIDATION CHANGES TO INCREASE DATA

Several aspects of the real-time quality control system were judged overly restrictive. The following improvements will be made resulting in additional data being transmitted:

- The consequences of a parity error will be made less inhibiting. Under the new system, only the measurement where the error is located is not transmitted. Previously, the entire message was not transmitted. One famous observation that never got transmitted because of this problem was from station 42020 during the formative stages of the March 1993 Superstorm (Gilhousen, 1994).
- The ability to transmit wind speeds from one anemometer and the wind directions from another was enabled. This circumstance can result from a cracked propeller or worn bearings on one anemometer and a faulty compass associated with

the other. At any given time in NDBC's network of 120 stations, this situation happens at one or two stations.

- The range check will be made less powerful because on a few, rare instances, it caused valid, but extreme, data to be withheld from transmission. More specifically, a measurement has to fail a time-continuity check for the previous observation before it can be deleted because of a range check.
- The real-time processing was installed on two UNIX workstations running simultaneously. This provides automatic backup capability without the need for manual intervention if a workstation fails or the processing on it crashes.

One other improvement concerns the timeliness of the validation. The validation is now done every 5 minutes instead of every 15 minutes. This means that NWS field offices are more likely to receive the observations before National Oceanic and Atmospheric Administration (NOAA) Weather Radio cut-off times.

## 6. CONCLUSION

Many changes were made recently to the real-time processing and quality control software for NDBC stations. These changes increase the overall amount of data while decreasing the chance of degraded measurements being transmitted. In recent years, the public began directly accessing NDBC observations from several home pages to verify marine forecasts and make their own marine safety decisions. One of these home pages receives more than three million "hits" a month! These Internet home pages obtain the reports from our encoded messages that undergo the quality control described. Therefore, automated validation techniques have become very important.

## 7. REFERENCES


Breaker, L.C., Gilhousen, D.B., and Burroughs, L.D., 1997: Preliminary results from long-term measurements of atmospheric moisture in the marine boundary layer, accepted by *J. Atmos. Oceanic Technol.*

Gilhousen, D.B., 1994: The value of NDBC observations during March 1993's "Storm of the Century", Weather and Forecasting, **9**, 255-264.

Gilhousen, D.B., 1988: Quality control of meteorological data from automated marine stations, Preprints, *Fourth Int. Conf. on Interactive Information and Processing Systems for Meteorology, Oceanography, and Hydrology*, Anaheim, CA, Amer. Meteor. Soc., 248-253.

Gilhousen, D.B., 1987: A field evaluation of NDBC moored buoy winds, *J. Atmos. Oceanic Technol.*, **4**, 94-104.

Meindl, E.A., and Hamilton, G.D., 1992: Programs of the National Data Buoy Center, *Bull. Amer. Meteor. Soc.*, **4**, 984-993.

National Data Buoy Center, 1996: Handbook of automated data quality control checks and procedures of the National Data Buoy Center.


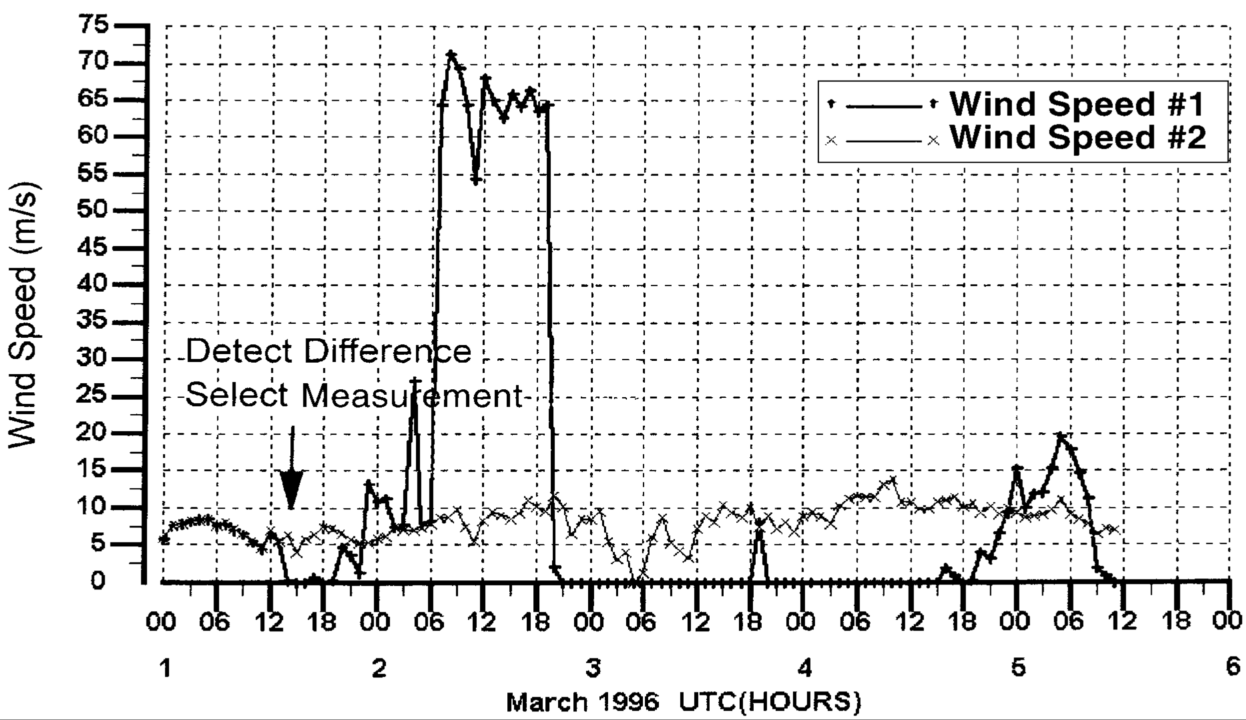


Figure 1. Time-series plot of wind speed measurements from two anemometers at buoy station 51003 located west of Hawaii