5.1 Use of the Consensus Reference Concept for Testing Radiosondes

Joe Facundo and Jim Fitzgibbon, Office of Operational Systems,

Silver Spring, Maryland and Sterling, Virginia

1. INTRODUCTION

The U. S. has been testing radiosonde manufacturer's radiosondes for many decades at test facilities around the country and has developed a number of test techniques for verifying performance. Recent advances in measuring the upper air atmosphere utilizing state-of-the-art referencing technologies and the development of new test techniques within the U.S. are now available for evaluating radiosonde performance to meet the more stringent climate monitoring requirements as well as for meteorological applications.

Today, numerical weather prediction (NWP) models can produce very good outputs of the meteorological and climatic parameters sensed by the diverse technologies described in this paper. As a result, comparing the sensor data with NWP statistics can be a very good method for assessing "consensus," i.e., how good does the data fit the model output? The goal is then to tie these technologies together with sound mathematical and statistical techniques.

The following is a concise definition along with a corollary definition:

- *Consensus Reference Testing: A preponderance of evidence derived from a suite of technologies and numerical weather prediction models converging on a statistical and repeatable set of acceptable thresholds. Determining the degree of consensus vs. non-consensus serves as the foundation for this concept.*

- *Corollary: No one technology or NWP model is viewed as absolute <u>truth</u> since all technologies have error characteristics; rather, each contributes some facet to a more <u>complete</u> understanding of the atmospheric variable or parameter under review.*

Each reference technology can play an important role in the *Consensus Reference Concept;* whereby, statistical techniques would be applied to the time-based and pressure/height radiosonde measurements of temperature, moisture variables, cloud bases, and winds as compared with the candidate reference or references.

For example, the climate community could converge on a set of criteria for performing these inter- and intra-comparisons and determine the minimal variance allowed to be compliant with the set of references. This approach would provide the climate community a wealth of knowledge in the contribution of each technology towards climate monitoring. Furthermore, the meteorological and climate communities can come to a *consensus* on which candidates meet the stated requirements and where more work is needed if they fall short.

Another goal would be to standardize and gain community acceptance of the testing methods, techniques, calculations, and data processing and to catalogue each one into a compendium for acquiring the "preponderance of evidence" as stated in the definition.

Finally, it should be noted that the techniques being discussed can be applied to vertical atmospheric measurements from other types of technologies, e.g., satellite or aircraft, in addition to radiosondes.

2. REFERENCE TECHNOLOGIES

There are numerous technologies which have the potential for meeting the goals of consensus referencing. Examples of these reference technologies include: Radiosonde Intercomparisons, NASA's Advanced Temperature Measuring system, RAMAN LIDARs for measuring the mid-to-upper tropospheric moisture, chilled mirror and LASER-diode humidity measurements, high-precision GPS measurements of height, Integrated Precipitable Water sensor using GPS techniques, various water vapor radiometers, and ground-based surface instrumentation for measuring surface parameters, radiation, clouds and weather. The following describe some of these technologies, although the entire suite above is considered candidates for this concept. As new technologies emerge in measuring the atmosphere accurately, they can be included into the concept.

2.1 Surface Observational Equipment

Detailed surface observations are used to compare surface conditions with the upper air measurements. Either the Automated Surface Observing System (ASOS) or the Radiosonde Surface Observing Instrumentation System (RSOIS) can be used to perform this function, since basic surface parameters such as the following are available:

Sky Condition
Visibility
Present Weather (type and intensity)
Freezing Rain (If Installed)
Thunderstorm (If Available)
Obscurations
Ambient Temperature, Relative Humidity, and Dew point Temperature
Wind (speed, direction, gusts, and direction variability)
Pressure (altimeter, station, density altitude, pressure altitude, and sea level)
Precipitation Amount
Long wave/Shortwave/Net Radiation

2.2 Functional Precision/Comparison

To have a good statistical sample the NWS will usually fly forty to fifty dual radiosonde flights for a functional precision or comparison test. "ASTM Standard, E 177, Standard Practice for Use of the Terms Precision and Bias in Test Methods" and "ASTM Standard, D 4430, Standard practice for Determining the Operational Comparability of Meteorological Measurements" covers these types of testing.

Functional Precision testing provides insight into the production quality of radiosondes, i.e., repeatability of the in-situ measurement, and is usually carried out in different meteorological and climatic regimes. Once testing is completed at all test sites and the data processed, the data is then combined to have a more definitive answer of the functional precision over the spectrum of weather and climate conditions. These then become a piece of the consensus dataset.

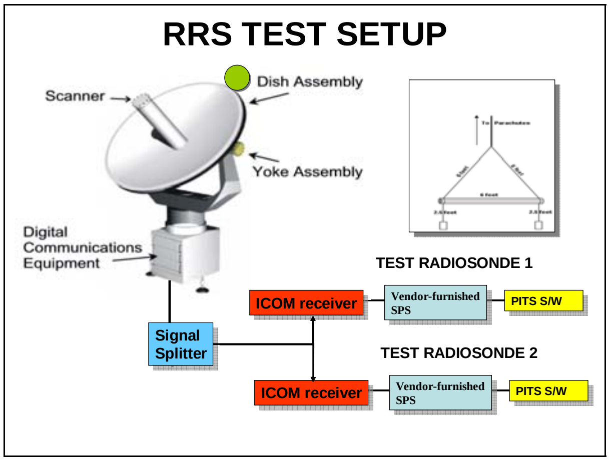

Figure 1. A possible test configuration.

The functional comparability test, which is an inter-comparison between two different types of radiosondes or other technologies, is performed in order to determine biases between the different types. For this test each candidate may be tracked using similar or different ground systems. Figure 1 illustrates one type of ground system setup. These flights should be conducted at synoptic times for best results since they can be "ensembled" with NWP derived data as well to develop a *consensus* view. Refer to x below.

2.2 Accurate Temperature Measuring (ATM)

This test is conducted to evaluate the accuracy of radiosonde temperature measurements and to validate any new or modified solar radiation correction algorithms being used with the radiosonde.

The ATM radiosonde has 5 temperature sensors; two white, two silver and one black. For the solution, only one of each color sensor was used. For each of the three different colored sensors the emissivity and absorptivity of the coatings have been pre-determined. This information is then used to solve simultaneous equations to determine the true temperature. This process eliminates the effects of the solar radiation and provides a true temperature measurement. This true temperature is then compared against the test radiosonde or devise that was flown on the same balloon.

An example of the ATM along with the test radiosonde is shown in Figure 2. Note, the multiple plots shown in this figure depicts the different outputs from the black, white and graycolored temperature sensors along with the test radiosonde. Several more plots are derived along with difference plots to complete the solution. Then a composite difference plot by time is generated indicating the expected difference between the two systems throughout the atmosphere. In this way, users of the data can understand the amount of solar radiation correction being applied to the actual temperature measurement.

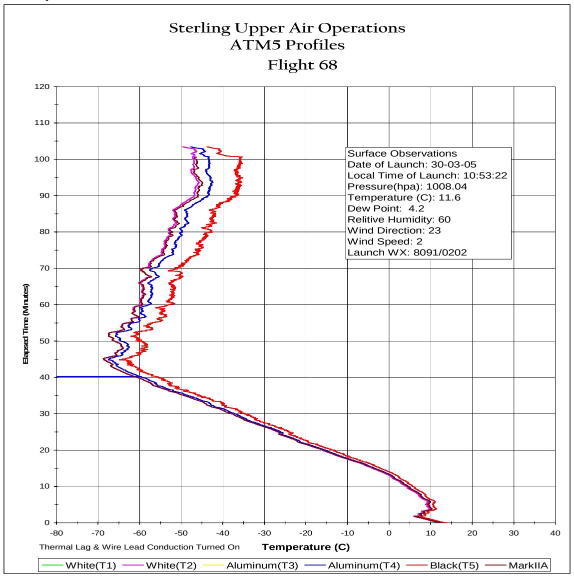

Figure 2. Example of ATM comparison flight.

2.3 LIDAR Inter-comparisons

Another exciting area being developed for this concept is with respect to inter-comparisons with RAMAN LIDAR measurements. The NWS through its NOAA-Howard University (HU) Center for Atmospheric Science (NCAS) agreement with HU is leveraging their RAMAN LIDAR situated at Beltsville, Maryland in partnership with NASA. A picture of the LIDAR with the one of the authors of this paper in the background is shown in Figure 3. RAMAN LIDAR offers the opportunity to depict in high-resolution the water vapor expressed as mixing ratio values to 20 km, the cloud bases, and rapid water vapor profile updates.

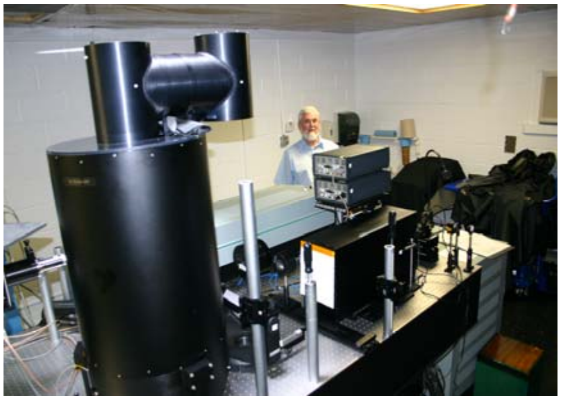

Figure 3. Example of Howard University LIDAR used for comparison.

Initial consensus reference testing was conducted this past summer during NASA's Water Vapor Validation Experiment – Satellite/Sondes (WAVES_2006) project. In addition, techniques developed by the Atmospheric Radiation Measurement Program in this area will be reviewed for their application to this approach.

3. NWP MODEL COMPARISONS

Where these techniques hold most promise is with comparing sounding data from diverse systems which are not all in the same in-situ space, e.g., radiosonde, aircraft, and satellites measuring different space and time domains with NWP model output fields. In essence, consensus referencing is akin to ensemble NWP forecasting whereby the consensus reference techniques would be analogous to various model outputs being compared for consistency.

3.1 Methodology

For example, Figure 4 reproduced from NOAA's Global Systems Division URL: http://gpsmet.noaa.gov/ruc20/ncep/06276/062761400_ncep_ne.html, depicts an integrated precipitable water (IPW) field color coded for the total amount. In the consensus reference testing approach, different technologies not physically co-located could be referenced against this analysis and statistics generated from their differences. The individual sensors would be compared against the background field. An aircraft flying from an airport could also have its IPW computed in a similar manner so that it would be aligned with the background field from the model.

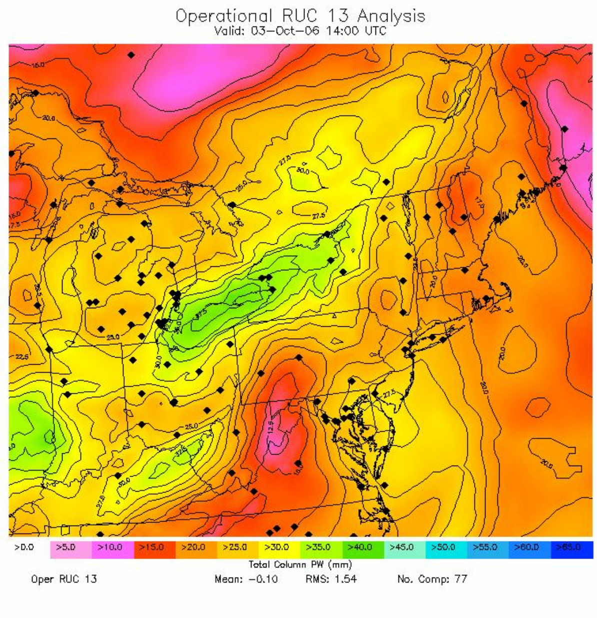

Figure 4. Example of RUC/GPS-IPW comparison product.

3.2 Aviation Model Output

Another example of consensus referencing is between the aviation model created by NCEP and test instrumentation as illustrated below. In this case differences in height from different technologies could be compared against various heights and consensus reference thresholds – horizontal bars – that would define the degree of consensus and non-consensus. These could then be determined all-year-round at different locations for radiosondes, satellites, aircraft, etc.

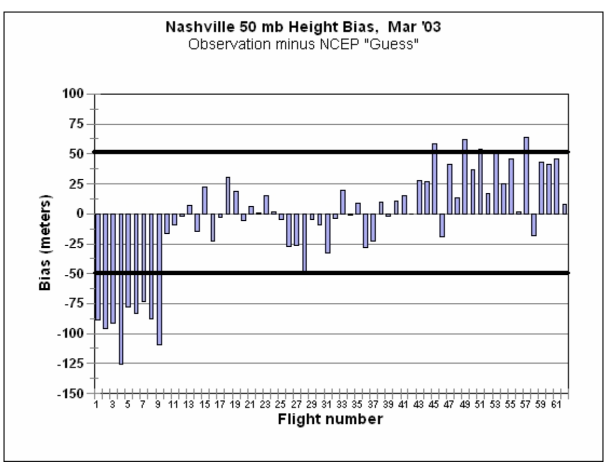

Figure 5. Example of Aviation Model/ Radiosonde comparison.

4. EXAMPLE OF CONSENSUS REFERENCE

The IPW-GPS sensor developed by the Office of Atmospheric Research within NOAA has great application for consensus referencing and serves as a good example of the technique.

4.1 Methodology

Figure 6 illustrates a time series of IPW measurements from the sensor and calculated IPW from the radiosonde. Thick, curvy lines denotes general trend of IPW data and dashed lines show approximate domain of all the data. Circles illustrate match-up of the radiosonde/IPW although not all match-ups are highlighted. Note, this approach could be accomplished for any measuring system providing vertical moisture profiles.

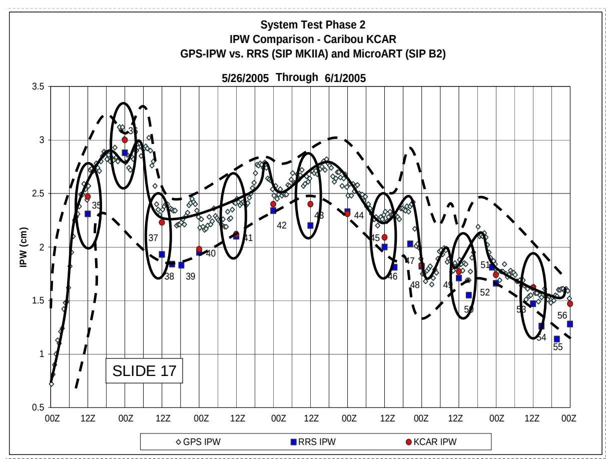

Figure 6. Example of Radiosonde/GPS-IPW comparison.

4.2 Data Analysis

In this form the data does not appear to have “information content” on how well the two match up. To obtain this level, the data needs to be re-organized into statistical graphs with consensus thresholds to ascertain the degree of consensus and non-consensus as shown in Figure 7. Note the thresholds do not have to be linear as is illustrated, but can be any shape representing the requirements for consensus and non-consensus. Rigorous statistics and mathematical treatment will be applied as the techniques are developed and introduced to the community for acceptance.

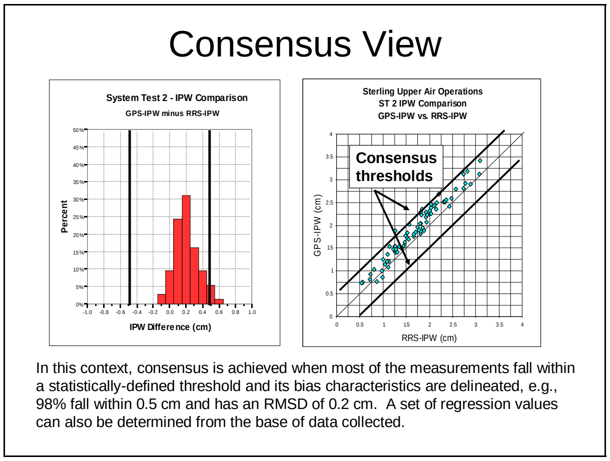

Figure 7. Example of consensus view of same data

5. ONE TECHNIQUE UNDER DEVELOPMENT

One technique under development by NWS with support from HU under the NCAS program is the development of the consensus referencing techniques for evaluating moisture parameters in the vertical. With the use of their LIDAR and the IPW-GPS, there is the potential for referencing different types of measurements each sampling a domain of atmosphere, but not necessarily identical space or time domains, and still be able to ascertain the degree of consensus among the different platforms. This will be further developed in a later paper.

6. Conclusions

The purpose of this paper is to inform the meteorological and climate communities about the potential for a consensus reference test concept, whereby an ensemble of tests are conducted and the results standardized to formulate a consistent pattern for evaluating upper air instrumentation and systems. Once the tests discussed in this paper are developed and proven, the plan is to document them into a catalogue for use by the wider community and conjoin them with a standard test process.

7. References

World Meteorological Organization 1996: Guide to Meteorological Instruments and Methods of observation. Sixth Edition WMO-No. 8, Geneva.

Schmidlin,F.J., Luers, J.K., and Huffman, P.D. 1986: Preliminary Estimates of Radiosonde Thermistor Errors. NASA Technical Paper 2637. Wallops Island, Virginia

NOAA Technical Report, NWS 44: Functional Precision of National Weather Service Upper-Air Measurements using VIZ Manufacturing Co. “A” radiosonde (Model 1492-510)

ASTM Standard, E 177, Standard Practice for Use of the Terms Precision and Bias in Test Methods.

ASTM Standard, D 4430, Standard practice for Determining the Operational Comparability of Meteorological Measurements.

AMS Extended Abstract, Testing Radiosonde Replacement System (RRS) Radiosondes – Part 1, Jim Fitzgibbon, and Joe Facundo, Office of Operational Systems, Silver Spring, Maryland

AMS Extended Abstract, Testing Radiosonde Replacement System (RRS) Radiosondes – Part 2, Jim Fitzgibbon, and Joe Facundo, Office of Operational Systems, Silver Spring, Maryland