4.4 Update on the Consensus Reference Concept for Testing Radiosondes

Joe Facundo, Carl Bower, 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. Examples of these reference technologies include: NASA's Advanced Temperature Measuring system, Howard University Atmospheric Observatory (HUAO) LIDARs for measuring the mid-to-upper tropospheric moisture, Snow White, high-precision GPS measurements of height, the Integrated Precipitable Water sensor using GPS techniques, various radiometers, and ground-based surface instrumentation to measure clouds and weather. Each reference technology can play an important role in the *Consensus Reference Concept;* whereby, data are integrated into information bases from which statistical techniques would be applied to the time-based and pressure/height candidate instrument measurements of say, temperature, moisture variables, cloud bases, and winds as compared to the references in use.

The climate and numerical weather prediction communities would assist in the preparation of the set of criteria for performing these inter- and intra-comparisons and determine the minimal variance allowed to be compliant with each reference, i.e., within consensus. This approach will provide these communities a wealth of knowledge in their overall performance, which can be repeated as often as necessary, and can serve as a methodology for allowing each community to come to a *consensus on which candidates meet the stated requirements and where more work is needed if they fall short.*

2. TYPES OF REFERENCE TESTS

Reference tests can be divided into several groups as illustrated in Figure 1. The test process is structured into layers beginning with Laboratory/Environmental Chamber Reference Tests followed by the Surface/In-situ Reference Tests and then Remote Sensing and NWP Model Comparison Tests. The latter four sets of tests can all be conducted simultaneously, thus reducing the test time and producing a rich data set for each test run. Tests will be further defined by such attributes as parameter, e.g, temperature, pressure/heights, moisture, and winds, meteorological and climatic events. The results from each test will be compiled into a Consensus Report documenting the entire test and findings both for individual comparisons and for the grouped data providing trends across meteorological and climatological boundaries.

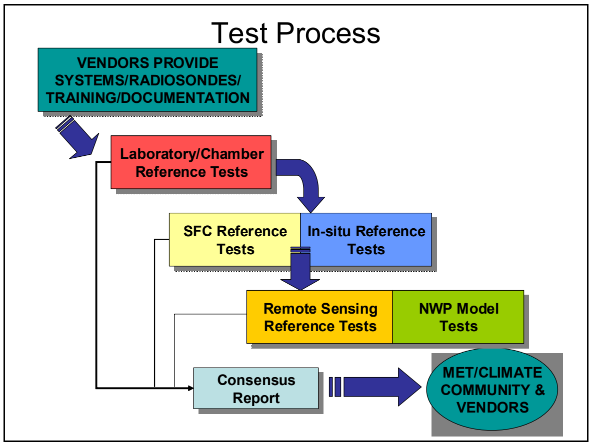


Figure 1. General test processes during Consensus Referencing.

2.1 Consensus Reference System

The Consensus Reference System consists of the following components:

- One or more ground systems for tracking radiosondes and reference instruments

- HUAO 30-meter tower for low-level measurements
- Surface systems
- Remote systems including GPS-IPW, Wind profiler, and radiometers
- Precision Digital Barometer
- Data Base Management System/Display

An *In-Situ GPS Reference* is also being pursued for independent measurements/calculations of geometric heights, geo-potential heights, derived-pressures, and the u- and v-components for calculating winds aloft. The plan is to integrate all these components at HUAO in Beltsville, MD as a joint facility with the university. Refer to Figure 2.

The real challenge with CRS will be integrating datasets from the diversity of technologies and synchronizing them within the frames-of-reference as discussed in Section 3.

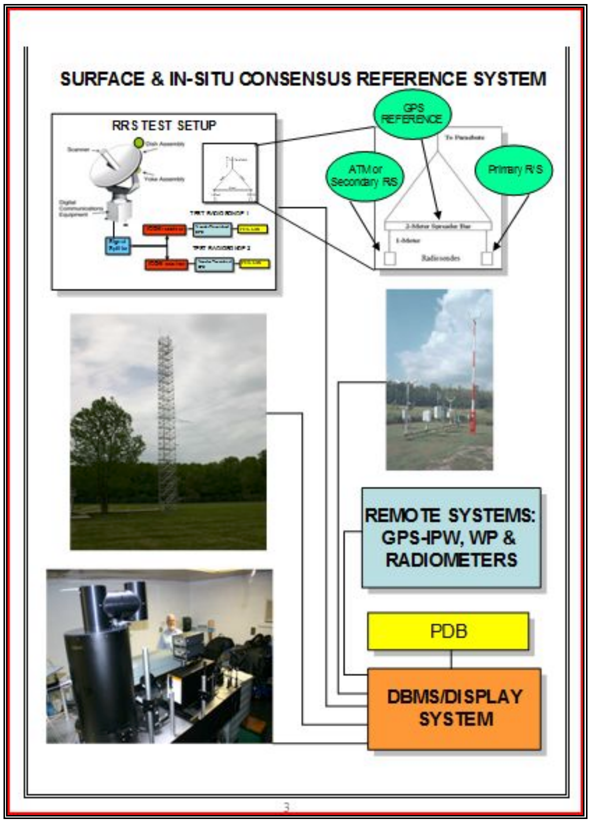


Figure 2. Consensus Reference System.

2.2 Temperature Referencing

Temperature referencing is centered on the use of the NASA ATM (see References for details) and if available, a temperature LIDAR.

.

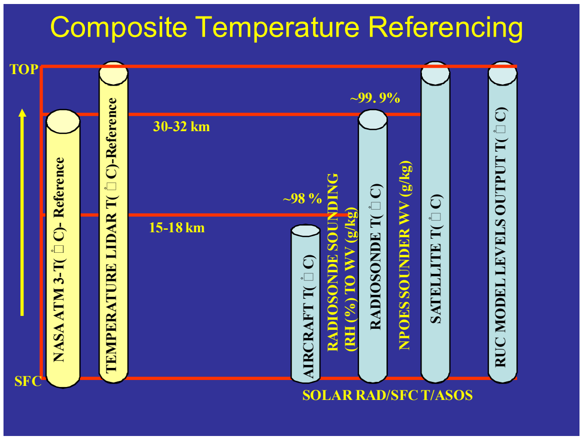


Figure 3. Temperature referencing systems and candidate test systems.

Figure 3 illustrates how the temperature referencing systems and candidate test systems would interface within this concept.

2.3 Pressure/Height Referencing

To evaluate pressure/height measurements aloft, the plan is to utilize independent GPS signals from a GPS reference devise suspended from the balloon and interfaced into the radiosonde for modulating the signals to the ground system.

2.4 Moisture Referencing

Moisture referencing is the most challenging of the measurements since water vapor is very variable. As figure 4 illustrates, one method is to utilize the all weather aspects of the GPS-IPW derived measurements as a reference. Other technologies, e.g., LIDAR, also provide excellent reference measurements, but are limited in one way or another and thus can only be used in a limited fashion within the consensus referencing concept.

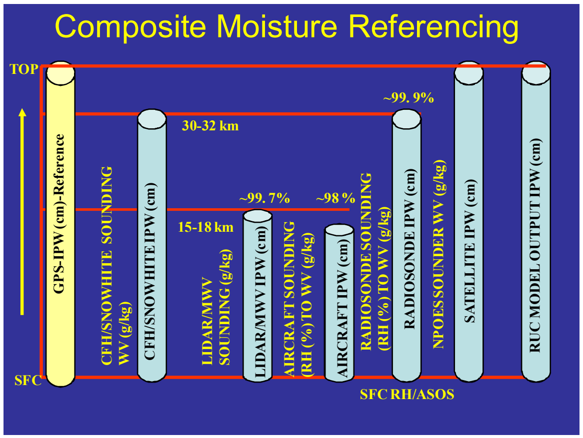
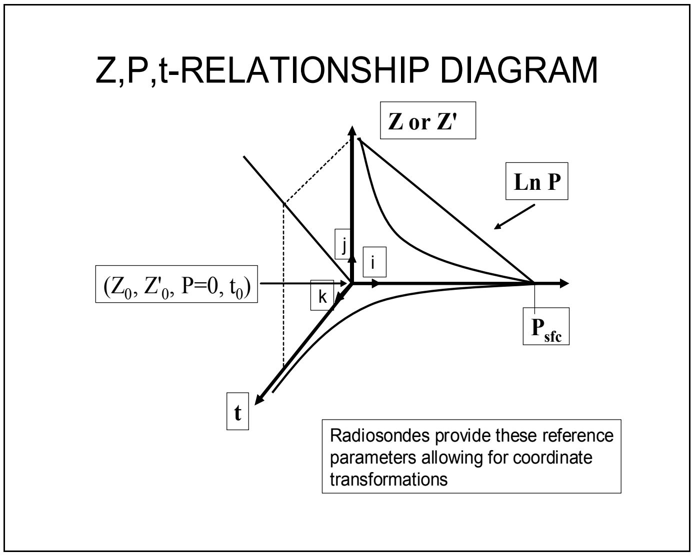


Figure 4. Moisture Referencing System and candidate test systems.

3. The Referencing Principle

The referencing principle is centered on *frames-of-reference* used by each component of the Consensus Reference System (CRS) as described in Figure 2. Theses frames-of-reference are based on the height (Z), pressure (P), and time (t) relationship that exists between these components. Heights can be expressed as geometric (Z) or geo-potential (Z') and be defined as Above-Ground-Level (AGL) or Mean-Sea-Level (MSL). For example, for those systems in Figure 2 measuring the upper air, one may be referenced using Z_{AGL} and P and another t and Z'_{MSL}, in which case a linkage can result with Z as the common denominator after translation. Radiosondes provide all four including P, t, Z and Z' (refer to Figure 5); with Global Positioning Satellite (GPS)-derived time and heights providing a very high degree of accuracy to each reference. Furthermore knowing Z or Z' one can also deduce P from the hypsometric equation or if it was GPS-derived, then directly if the surface pressure is known, independently. This concept allows for an excellent method for transforming data sets between CRS components. One can also compute the ln(P) to derive a relatively straight line instead of an exponential curve.



Radiosondes provide these reference parameters allowing for coordinate transformations

Figure 5. Frames-of-Reference.

One future application of the referencing principle is the ability to compare measurements from different systems, e.g., balloon, satellite, RAMAN LIDAR, aircraft, all moving in different time and space domains. This is because one could compare each of these against a numerical weather prediction (NWP) model output, say of Integrated Precipitable Water (IPW), and determine if equivalent measurements are being produced by each even if they are not in-situ or measured at exactly the same time. For this reason, consensus referencing can become a very powerful tool for assessing observing platforms whether they are remote or in-situ or determine the adequacy of NWP models to represent atmospheric parameters.

3.1 Segmented Layers

Another aspect of the referencing principle is to segment the atmosphere into its component layers:

- Surface
- Planetary boundary layer (PBL)
- Troposphere/Tropopause
- Stratosphere

Testing will, therefore, be segmented into these atmospherics blocks to ensure each one is being characterized properly for its specific meteorological and climatological aspects.

3.2 Consensus Reference Thresholds

The concept of *Consensus Reference Thresholds (CRTs)* is central to understanding the consensus referencing concept. To meet the premise as defined in Section 1, one needs to know how close the measurements will need to agree over what environmental ranges, and over what seasonal and local components (see Figure 6). This is accomplished by establishing CRTs for each referenced dataset being compared using one of several methods as follows:

- Empirical Method – based on the professional expertise of scientists who understand the science of measurements and what the degree of agreement needs to be for consensus.
- Statistical Methods – based on established statistical methods, e.g., confidence limits or intervals, resulting from dataset analyses.
- Mathematical Methods -- based on established mathematical methods, e.g., allowable percent offsets, resulting from dataset analyses.

CRTs can also be characterized into zones of agreement for developing a further understanding as to the degree of consensus within a dataset. These are divided into three zones as follows:

- Strong Consensus (SC) – bounded zone defining best performance characteristics over the expected range of values.
- Marginal Consensus (MC) – bounded zone defining minimal performance characteristics over the expected range of values.
- No Consensus (NC) – unbounded region defining poor-to-no performance characteristics over the expected range of values.

An example of the degrees of consensus is shown in Figure 6. In this figure one sees that there are SC, MC, and NC regions, called threshold zones. A mathematical – could be statistical as well – technique might be applied to further understand the degree of consensus as follows:

To compute the degree of consensus, calculate the following:

$$C'_{SC} = \frac{N\,(T_{SC})}{N},$$

$$C'_{MC} = \frac{N\,(T_{MC})}{N},$$

$$C'_{NC} = \frac{N\,(T_{NC})}{N},$$

(Where $N\,(T_x)$ are the number of samples within each threshold zone, and N is the total sample size.)

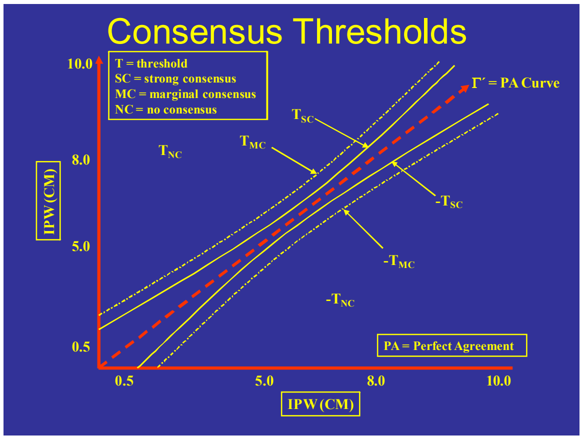


Figure 6. Consensus Thresholds.

For consensus to be evident, the expected *degrees of consensus* could be as follows:

EXAMPLE

	Thresholds	Actuals
	$97\% \le C'_{SC} \le 100\%$	98 %
	and	
C' =	$0\ \% \le C'_{MC} \le 3.0\%$	1.3 %
	and	
	$0\ \% \le C'_{NC} \le 0.5\%$	0.7%

(Where C'_{SC}, $C'_{MC,}$ C'NC are the percentage of samples within each threshold zones for the SC, MC, and NC, respectively). This means all three degrees would have to be satisfied within the constraints defined above in order for consensus to be achieved. If even one category under the Actuals column is not met, then consensus is not achieved and further analysis would define where exactly the consensus fails. Further work would be needed in this area to eventually result in consensus. One could also portray these ranges using statistics computing, say, standard deviations (σ) as follows:

$+/-1\sigma \leq C'_{SC}$	1st degree of consensus
$+/-2\sigma \leq C'_{MC}$	2nd degree of consensus
$2\sigma > C'_{NC}$	3rd degree of consensus

4. 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 within a standard test process.

5. Acknowledgements

The NWS wishes to acknowledge the efforts of those involved with the HUAO WAVES 2007 project. Special appreciation goes to Dr. Everett Joseph and Demetrius Venable along with Michael Hicks is in order. Within NWS, the SAIC support contractor, especially Ashby Hawse and Michael Baldwin deserve special attention. Finally, special thanks are in order for Seth Gutman from NOAA/OAR, Belay B. Demoz (NASA/Goddard), and Frank Schmidlin (NASA/Wallops Island) who have provided inspiration and a deeper understanding towards developing these concepts.

6. 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

AMS Extended Abstract, Use of the Consensus Reference Concept for Testing Radiosondes, Joe Facundo and Jim Fitzgibbon, Office of Operational Systems, Silver Spring, Maryland and Sterling, Virginia

The Consensus Reference Methodology as it Applies to a Radiosonde under Test, Ryan Brown, Joe Facundo and Jim Fitzgibbon, Office of Operational Systems, Silver Spring, Maryland and Sterling, Virginia