A Comparison of 12 US Liquid Hydrocarbon Flow Standards and the Transition to Safer Calibration Liquids

John Wright, Aaron Johnson, Gina Kline,
Chris Crowley, Jodie Pope, Vern Bean
National Institute of Standards and Technology

Kevin John, Jeremy Latsko
Air Force Metrology and Calibration

David Pfeffer
Air Force Primary Standards Lab

James Winchester
Arnold Air Force Base and University of Tennessee Space Institute

Michael Case, Gary Mahoney, Gary Quigley
Hill Air Force Base

Jason Bellavance
Robins Air Force Base

Jon Fitzgerald, Bradley Ramsey
Tinker Air Force Base

Wes England, Russell Kauffman, Michael Vickers
Army Primary Standards Lab

Dennis Page
US Army TMDE Support Region-West, Engineering, White Sands Missile Range

Juan Garcia, William Monette, Curtis Van Diepen
TMDE Support Center, Corpus Christi Army Depot

Timothy Tippins, Jeffrey Fetzer
Navy Mid-Atlantic Regional Calibration Center

Michael Wusterbarth, Andrew Yee
Badger Meter, Inc. / Flow Dynamics, Inc.

Mark Evans, Rusty Gavagan, Jeff Morrissey
Flow Technology, Inc.

During 2010 and 2011, NIST piloted a 12-laboratory comparison of hydrocarbon liquid flow calibration standards spanning the range 3.8 L/min to 38 L/min. The laboratories were in mutual agreement within the expected 0.3% uncertainty, which is approximately half as large as the differences measured in a similar 1988 comparison. The transfer standard (a pair of turbine flow meters in series) introduced an uncertainty of 0.17%* into the comparison. The comparison protocol used methods that were developed during international comparisons including: using uncertainty weighting to generate a best fit comparison reference curve, using statistical criteria to remove discrepant results from the fit, assessing and including in the data analysis the uncertainty contributed by the transfer standard, and reporting a standardized degree of equivalence between the participants. Several laboratories used mixtures of propylene glycol and water (PG + W) instead of Stoddard solvent (the commonly used surrogate for jet fuel) because the PG+W mixtures are safer and cheaper to manage environmentally. This comparison and other studies show that there is no significant difference in the calibration results between Stoddard solvent and a PG + W mixture with the same kinematic viscosity. Therefore, NIST is changing its calibration fluid to PG + W and encourages other laboratories to do the same.

---

* All uncertainties are approximately 95% confidence level (*k*=2) unless otherwise stated.

Introduction

The aerospace industry and the US Department of Defense measure fuel consumption in jet engine test stands and in other applications using turbine meters that are periodically calibrated against reference standards. The end users require flow measurements with uncertainties < 0.5% and the test uncertainty ratio of 3 to 1 through multiple calibration layers imposes low uncertainty demands on the reference standards (< 0.1%). Calibration laboratories maintain their traceability, proficiency, and understanding of the behavior of the flow meters over wide ranging operating conditions so that they can assist the end users in correctly using flow meters to achieve their required uncertainties. Some labs conduct periodic comparisons with NIST by sending their check standards (usually turbine meters) to NIST for calibration. This is a valuable element of quality assurance, but it does not compare secondary or tertiary labs directly. Hence it is also valuable to periodically send a transfer standard to a large number of participants to confirm that the traceability hierarchy is functioning well and that there is the desired degree of equivalence between calibrations from all of the labs. The last time such a comparison was conducted for this sector was 1988 [1].

During the last decade, national metrology institutes (NMIs) such as NIST conducted key comparisons to demonstrate that they meet their uncertainty claims. Flow key comparisons are organized by the Bureau International des Poids et Mesures and its NMI members in the Working Group for Fluid Flow (WGFF). These international comparisons have developed a body of knowledge and consensus about comparison methodology. Some examples are:

- Publication of the Guidelines for CIPM Key Comparisons [2] which formalized aspects of planning and conducting a comparison such as: the selection and number of participants and using an agreed upon protocol that covers (1) the schedule, (2) instructions for operating the transfer standard, (3) reporting results (including uncertainty), and (4) communication issues such as keeping results confidential, resolution of anomalous results, and the review of draft reports before dissemination.
- Papers on processing comparison data, particularly Cox's The Evaluation of Key Comparison Data [3], which documents the calculation of the key comparison reference value (KCRV), its uncertainty, and degrees of equivalence (differences) between laboratories. Cox recommends two methods for calculating the KCRV: 1) an "uncertainty weighted average"* of the participants' results or 2) in case of statistically discrepant results (outliers), the median.
- The Working Group for Fluid Flow developed specific recommendations for comparisons that use flow meters as transfer standards. They include thorough preliminary testing of the transfer standard to determine its sensitivities to environmental and installation variables and the uncertainty it contributes to the comparison [4]. If a transfer standard is sensitive to the fluid temperature (or other variables), this must be identified and quantified before the comparison begins. Otherwise, differences introduced by the sensitivity of the transfer standard to environmental and installation will be incorrectly interpreted as lab-to-lab differences. Another idea adopted by the WGFF is to fit a curve to the participants' flow meter calibration data and to calculate differences of each lab from this "comparison reference curve" [5].

The present comparison of labs that calibrate flow meters for hydrocarbon liquids intentionally exploited the experience gained during prior international comparisons. In the following sections we will describe the reference and transfer standards used, the working fluids, details of the comparison data analysis, and the results of the comparison.

The Reference and Transfer Standards

All of the comparison participants used volumetric piston provers as the reference flow standard, similar to the NIST 20 L piston prover shown in Figure 1 [6]. A variable speed motor pushes a piston through a cylinder of known diameter, driving the calibration fluid through connecting piping to the meter under test. The position of the piston (and hence the volume of liquid pushed out of the cylinder) is measured with optical encoders (20 μm resolution) that are attached to the piston shafts. Valves in the connecting piping are automatically switched to maintain flow in the positive direction through the meter under test even though the piston alternates direction when it reaches the cylinder ends. To avoid cavitation at the meter under test, the entire system is placed under pressure by an external gas source. Pressure and temperature sensors allow corrections for storage effects, such as changes in the mass of the liquid in the connecting piping due to changes in liquid density over time.

---

* Using the inverses of the squares of the participants' uncertainties as weights gives greater significance in the comparison reference value calculation to laboratories with lower uncertainty.

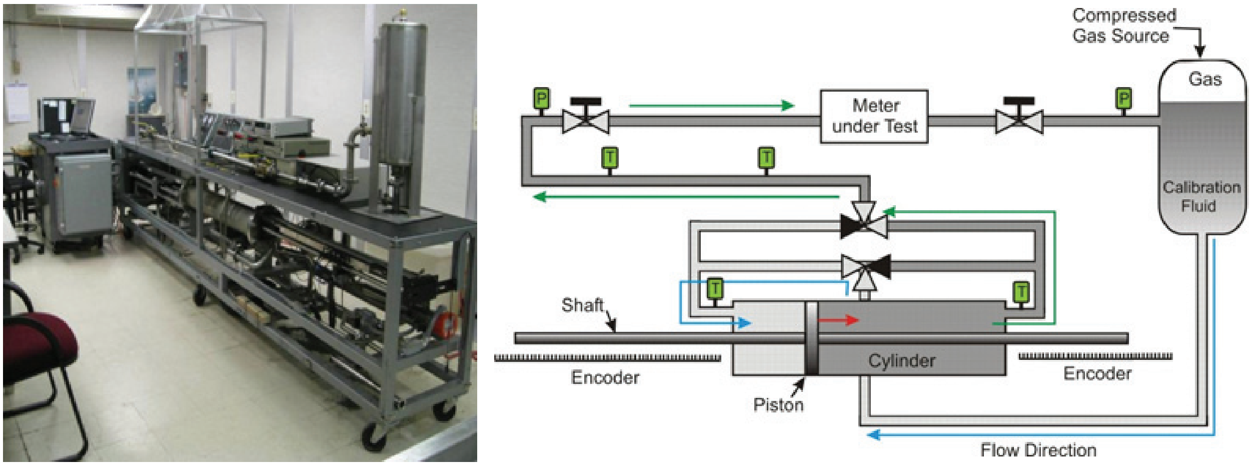


Figure 1. The NIST 20 L piston prover. Flow directions in the schematic are shown for the piston travelling from left to right. The T and P symbols represent temperature and pressure sensors.

The transfer standard for this comparison was two 1.3 cm nominal diameter dual rotor turbine meters installed in series. A plus-sign-shaped flow straightener was installed in the approach tube upstream of each turbine meter. Each participating laboratory recorded the sum of the dual rotor frequencies from the output of the manufacturer's signal conditioner. The protocol called for: (1) testing of the transfer standard at three nominal flows (3.8 L/min, 12 L/min, and 38 L/min), (2) reversing the order of the flow set points, and (3) reversing the order of the two meters (configurations 1 and 2). Steps (2) and (3) were used to evaluate hysteresis, installation effects, and reproducibility. Five data points were collected at each flow set point and meter configuration, resulting in 120 data points for each of the two turbine meters. Approximately halfway through the comparison, one of the turbines (SN 5644) was damaged by a piece of debris in the flow and it was replaced by a new turbine (SN 5852). One of the original two turbines (SN 5643) functioned well throughout the entire comparison.

Laboratory	Liquid	Kinematic Viscosity (mm2/s)	U (flow) or U(PS) (k = 2, %)
NIST	SS, PG + W	1.22	0.07
Air Force Primary Standards Lab	SS	1.29	0.05
Arnold Air Force Base	PG + W	1.2	0.05
Hill Air Force Base	SS	1.32	0.05
Robins Air Force Base	SS	1.24	0.05
Tinker Air Force Base	SS	1.3	0.05
Army Primary Standards Lab	SS	1.28	0.043
TMDE Support Center Corpus Christi Army Depot	SS	1.27	0.06
Navy Mid-Atlantic Regional Cal Center	SS	1.23	0.08
Flow Dynamics, Inc.	SS	1.31	0.04
Flow Technology, Inc.	SS	1.26	0.036
University of Tennessee Space Institute	PG + W	1.26	0.06 to 0.19

Table 1. Summary of comparison participants. Laboratory name, reference standard liquids (SS = Stoddard solvent, PG + W = propylene glycol and water mixture), its nominal kinematic viscosity at the temperature of the measurements, and the approximately 95% confidence level uncertainty for flow measurement reported by each participant are shown.

Calibration Liquids Used

Today, most laboratories that calibrate turbine meters for jet fuel applications use Stoddard solvent* as a surrogate fluid because it is less flammable and toxic than jet fuel, however, it still presents fire and biological hazards. Recently, Arnold Air Force Base Calibration Laboratories and NIST replaced Stoddard solvent with biologically and environmentally benign mixtures of propylene glycol and water. The Army Primary Standards Laboratory is also studying PG + W as an alternative calibration liquid [7]. A mixture of approximately 7% by weight (or volume) propylene glycol in water matches the kinematic viscosity of jet fuel at 21 °C (approximately 1.2×10^{-6} m^2/s) and pure propylene glycol has a kinematic viscosity of approximately 50×10^{-6} m^2/s, which matches the middle of the range of hydraulic oils at 21 °C. Propylene glycol ($C_3H_8O_2$) is commercially available, "generally recognized as safe" by the Food and Drug Administration, and is an ingredient in many consumer products such as cosmetics and food. Calibration laboratories that replace Stoddard solvent with mixtures of propylene glycol and water will (1) reduce inhalation danger to workers, (2) eliminate fire danger, and (3) decrease the cost of disposal of hydrocarbon liquids. NIST calibrations of turbine meters using flows of many mixtures of propylene glycol and water agreed with NIST's Stoddard solvent calibrations within 0.02% [8]. These results validated the theory for the dependence of turbine meter calibrations on the fluid properties density and viscosity. NIST's theory for turbine meters also incorporates the effects of bearing friction and fluid drag. In addition, it correlates data spanning a 200:1 flow range with liquid mixtures spanning a 42:1 kinematic viscosity range. Based on these results, the consensus of a NIST Workshop held in September 2011 was that where practical, NIST and other calibration laboratories should transition from Stoddard solvent to propylene glycol and water mixtures in their calibration services.

Data Processing

The calibration results were reported using the dimensionless Strouhal (St) and Roshko (Ro) numbers. In this comparison the Strouhal number was defined as:

$$St = \frac{\pi f D^3}{4Q_{\mathrm{MUT}}} \tag{1}$$

where Q_{MUT} is the actual volumetric flow at the meter under test (i.e. the reference flow measured by the participant), D = 1.3 cm is the diameter of the flow meter, and f is the sum of the two rotor frequencies from the meters under test. The Roshko number was defined as:

$$Ro = \frac{f D^2}{\nu} \tag{2}$$

where v is the value of liquid kinematic viscosity (i.e., density divided by dynamic viscosity) at the fluid's temperature and pressure. The diameter in Eqns. (1) and (2) was corrected for thermal expansion, but these corrections were insignificant (0.01% or less)**.

The three flow set points for the comparison were selected in the viscosity independent range of the turbine meters [8]. Preliminary testing at NIST identified the viscosity independent range of each turbine meter at $Ro > 5 \times 10^4$.

NIST used a best-fit polynomial to obtain comparison reference curves (CRCs) for each turbine meter. Only data for each turbine meter in the upstream position was used to obtain the CRCs and in most of the comparison analysis. The data were fitted using three different options (see Figure 2): (1) using equally weighted data from the participants (including NIST) at the three flow set points, with discrepant results removed, (2) using uncertainty-weighted data [$1/U^2(St)$ where $U(St)$ is the expanded uncertainty of the Strouhal number] from the participants at the three flow set points, with discrepant results removed and (3) using equally weighted data from NIST alone, including both preliminary and post-comparison testing to check the transfer standard stability [3]. (Note that the numerous data used to obtain the "NIST only" fit are not shown in Figure 2, only the averages from one run of the protocol that was used as the NIST comparison data.) The largest difference (0.08%) between the three versions of best fit curves occurred at the medium flow set point for the replacement turbine meter, SN 5852 (see Figure 2C insert). The uncertainty-weighted fit to all participants (option 2) was used as the comparison reference curve for all three turbine meters.

A polynomial in log (Ro) was used to fit the calibration curves:

$$St_{CRC} = a_0 + a_1 \log(Ro) + a_2 \log(Ro)^2 + a_4 \log(Ro)^4 \tag{3}$$

The coefficient a_4 was zero for two of the turbines (SN 5644 and 5643) but a nonzero value was necessary to obtain an acceptable fit to the data for the replacement turbine (SN 5852). Also, because of the unusual shape of the SN 5852 calibration curve, an extra set of NIST data (distributed over a wider range than the comparison flow set points) was added to the fitting process, shown as "NIST extra" in Figure 2C. This produced a better fit at the endpoints of the comparison flow range. Once the comparison reference curves for the three turbines were established, the results of the comparison were analyzed by examining the percent difference between each participant's results and the CRC.

---

[*] MIL-PRF-7024E Type II, a light mineral oil.

[**] Several presentation methods for turbine calibration data are available. Here we use Strouhal versus Roshko numbers, but the K factor = f/Q_{MUT} versus f/v is also commonly used as is the Reynolds number for the abscissa. All would have worked equally well in this comparison. Strouhal versus Roshko numbers have the advantage of being dimensionless and accounting for thermal expansion effects on the meter dimensions (not a significant effect here). While the Reynolds number is dimensionless, the Roshko number has the advantage that it does not require Q_{MUT} and hence avoids an iterative process when the calibrated turbine meter is used to measure flow.

Uncertainty of the Strouhal Number

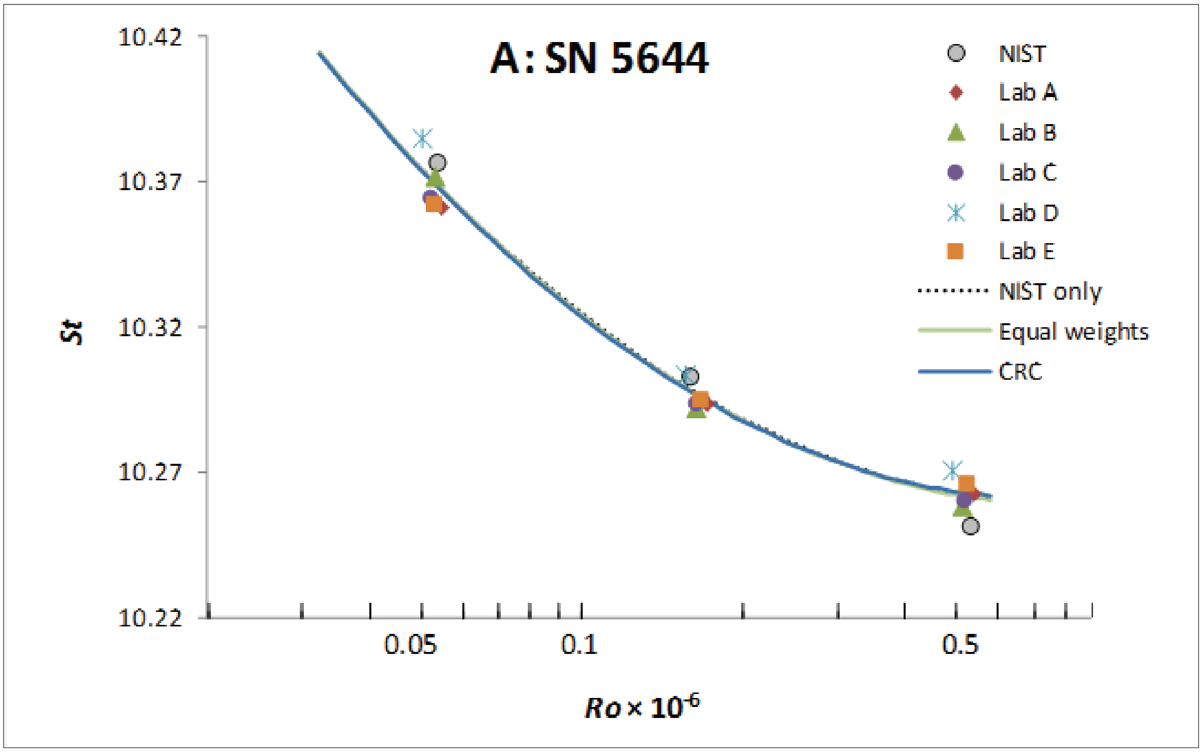
The uncertainty of the Strouhal number values reported by each laboratory are necessary inputs for the comparison data processing. They are used for performing chi-squared tests to eliminate outliers from the CRC calculations, for the weighting in calculating the CRC, and for calculating the uncertainty of the CRC [3]. In this comparison, the uncertainty components related to the transfer standard were similar in magnitude to the largest primary standard uncertainty, leading to a large and nearly constant value for the uncertainty of the Strouhal number for all participants (0.15% to 0.19%, $k = 2$). This resulted in approximately equal significance for each participant in the uncertainty-weighted best fits.

The expanded uncertainty of St was calculated using the formula:

$$U^2(St) = U^2(PS) + U^2(S) + U^2(v) + U^2(f) + U^2(T) \qquad (4)$$

where the variable names and how they were determined are:

1. $U(PS)$ is the expanded uncertainty of the participant's primary standard. These values were reported by each lab (see Table 1) and ranged between 0.036% and 0.19%. $U(PS)$ does not include the reproducibility of the meter calibration data from the comparison.
2. $U(S)$ represents the calibration stability of the transfer standard over the course of the comparison. This value was determined by doubling the standard deviation of the residuals of best fit curves for all NIST preliminary testing at the set point flows over the 2 year period of the comparison. $U(S)$ was 0.1%, 0.14%, and 0.1% for SN 5644, SN 5643, and SN 5852 respectively. We also studied the ratio of the upstream and downstream meter frequencies in order to check that the two turbines maintained calibration stability while in use at the participating labs and found similar variability. This component is also covers the uncertainty due to irreproducibility in the participant labs.
3. $U(v)$ is the expanded uncertainty in St due to kinematic viscosity effects. Participants used either Stoddard solvent or propylene glycol and water mixtures that varied in kinematic viscosity between 1.20 × 10^{-6} m^2/s and 1.32 × 10^{-6} m^2/s. A NIST evaluation of turbine meters used for Stoddard solvent and PG + W mixtures determined that the fluid change introduced an expanded uncertainty of 0.02% within the viscosity-independent range of a turbine meter [8]. This sub-component was root-sum-squared (RSS) with a second sub-component, resulting from uncertainty in the kinematic viscosity values used by each lab. This second sub-component



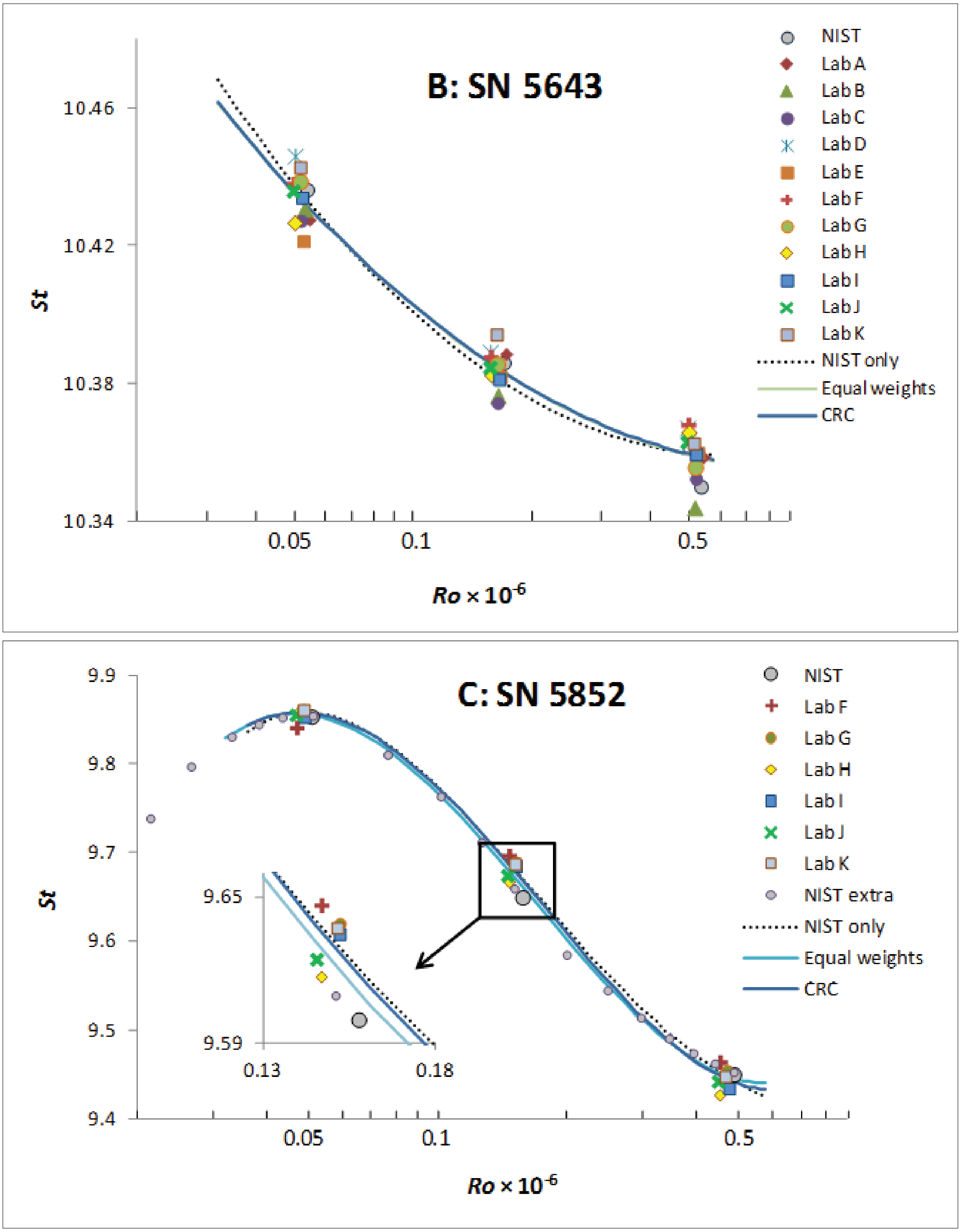


Figure 2. Best fit comparison reference curves for the 3 turbine meters and the data used to obtain them. Three versions of best fit curves are shown, but uncertainty-weighted fits to all participants were used as the comparison reference curve (CRC). Note that the "equal weights" curve is not visible in panels 2A and 2B because it is covered by the CRC. The insert in panel 2C shows two clusters of data separated by 0.3% at the medium flow.

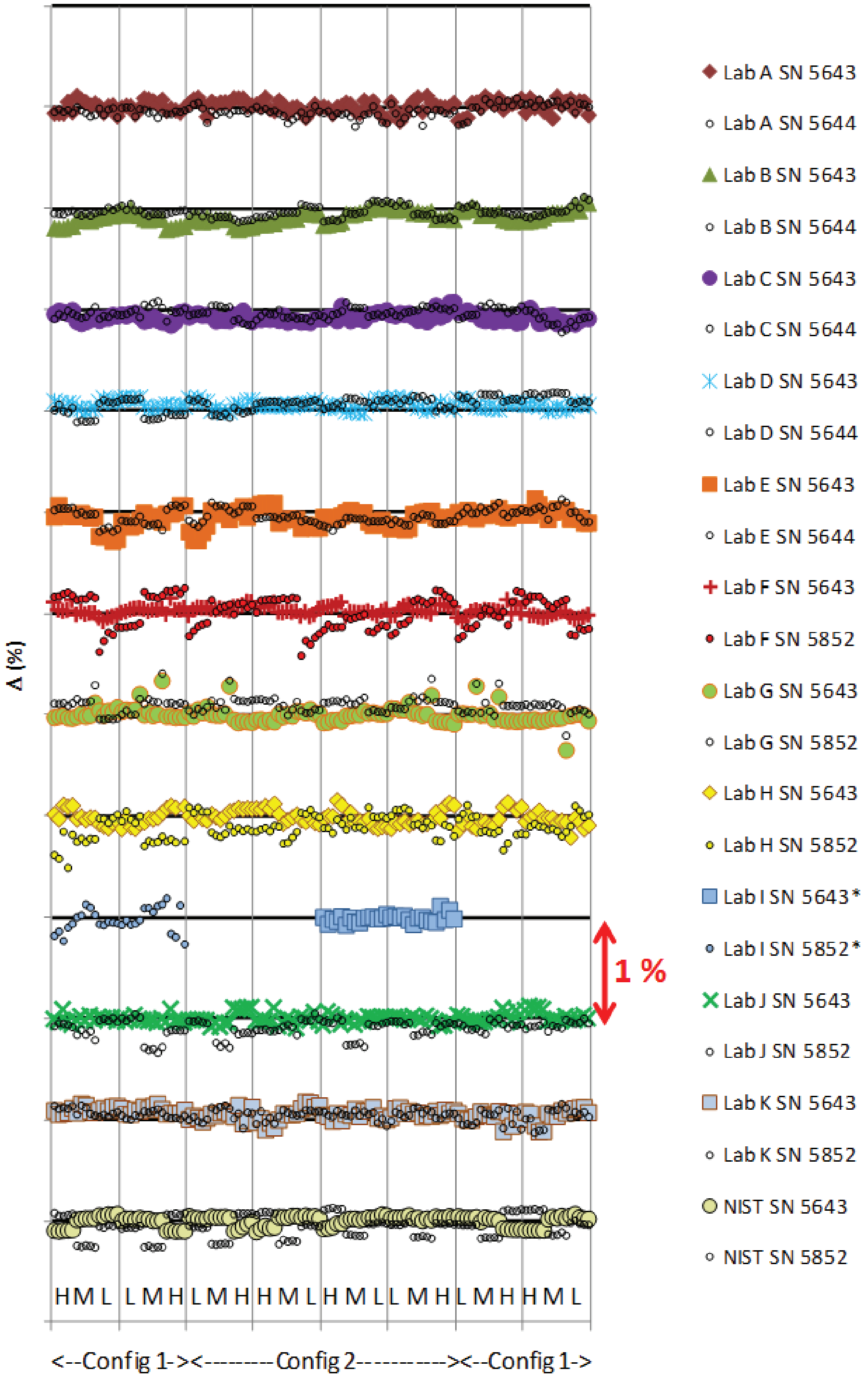


Figure 3. Data for both turbine meters collected by each participant for all configurations (upstream or downstream positions).
*The data from Lab I were not collected with 2 meters in series.

was determined by calculating the slope of the *St* vs. *Ro* plot for each meter at the 3 comparison flows and assuming a 4% expanded uncertainty in the reported v values due to either (1) temperature measurement uncertainty or (2) problems with the systems used to determine the relationship between v and T. The uncertainty in *St* introduced by a 4% uncertainty in v varied between a negligible value (where the calibration curves are flat) to 0.1 % for SN 5852 at the medium flow set point. Using the worst case of 0.1 % for the second component, the root-sum-square of the two $U(v)$ sub-components rounds to 0.1%.

4. $U(f)$ and $U(T)$ are the uncertainties due to frequency and temperature measurements respectively and both are negligible in this comparison. In a few cases, there were obvious problems in frequency measurements and these data points or sets were re-measured or removed. The effects of temperature uncertainties via thermal expansion corrections are negligible (<0.01% for 6 °C) and the effect of temperature on kinematic viscosity was included in $U(v)$ above.

Combining these components leads to an expanded, 95% confidence level U (St) of 0.15% to 0.19%, depending on the participant and turbine meter considered. Combining the components U (S) and U (v) gives an expanded uncertainty due to the transfer standard of 0.14% to 0.17%.

Comparison Results

Figure 3 shows the data processed for the comparison. The y-axis shows the difference from the CRC in percent (Δ), with each participant's results offset by an integer multiple of 1% so that the data for different labs do not overlap. The x-axis is a time series of the 120 individual points as measured in the protocol. The x-axis is labeled with both the configuration (1 or 2, i.e. which meter is in the upstream position) and with the flow set point (low=L, medium=M, and high=H). Configuration 1 is the arrangement with either SN 5644 or its replacement, SN 5852 in the upstream position. Configuration 2 is the arrangement with SN 5643 in the upstream position.

NIST testing at the conclusion of the comparison showed that the pressure drop through the transfer standard was large, and unless the meters were calibrated at pressures > 480 kPa, incorrect, low *St* values were measured at the downstream turbine, probably due to cavitation. The data affected by this problem was either removed or additional testing was done at higher pressures to remove this source of error from the comparison results. One participant re-tested because of interference between the two turbine meter outputs in their data acquisition system. Lab I could only test one meter at a time because of data acquisition limitations.

Figure 4 presents the comparison results as standardized degree of equivalence, E_n, in which the difference between each participant and the comparison reference curves (Δ) is normalized by the uncertainty expectations for the difference:

$$E_n = \frac{\Delta}{U(\Delta)} \tag{5}$$

where $U(\Delta)$ is the $k = 2$ uncertainty of the difference between a participant result and the CRC [3]. By this measure, E_n values between -1 and 1 indicate that a participant is in agreement with the comparison reference curve within uncertainty expectations. All points for all labs fall within this bound for SN 5643. Two labs have $|E_n| > 1$ for SN 5852. The figure uses data from Figure 3, for each turbine meter when it was tested in the upstream position, i.e. configuration 1 for SN 5644 or SN 5852 and configuration 2 for SN 5643. Each point in Figure 4 represents the average of the 20 individual data points at the low, medium, and high flow set points, labeled as, L, M, or H, respectively.

Discussion

One of the two turbine meters was damaged and replaced about half way through the comparison. Using data from the meter that worked throughout (SN 5643), all participants had $|E_n| < 1$ (within uncertainty expectations) for all 3 flow set points and the largest difference between any two labs was 0.27%; given the ability of the transfer standard and protocol to resolve differences, the participants met their uncertainty claims. For SN 5852, two labs had $|E_n| > 1$ and the largest difference between any two labs was 0.39%. Two of the three $|E_n| > 1$ points were due to results for SN 5852 at the medium flow falling in two clusters separated by 0.3% (see Figure 2C). Since these lab-to-lab differences are not found for the other meter (SN 5643), they can be attributed to SN 5852 and not the laboratory reference standards. In fact, there is a noticeable increase in the range of E_n values for the replacement turbine meter relative to the other two turbines (see Figure 4). The lab-to-lab differences measured in this comparison are approximately half as large as those measured in the 1988 comparison [1].

The ability to discern differences between the participating labs was hampered by uncertainty components related to the transfer standard: (1) long term calibration stability (0.1% to 0.14%) and (2) kinematic viscosity effects (0.1%) which are large compared to some of the uncertainties of the primary standards reported by the participants (0.036% to 0.19%). The long term calibration stability of the transfer standard was assessed using (1) repeated calibrations performed at NIST before, during, and after the comparison, (2) the variance of the meter output frequency ratios when tested in series by each participant, and (3) the range of points in the northeast to southwest direction in Youden plots [9]. All three approaches gave similar results,

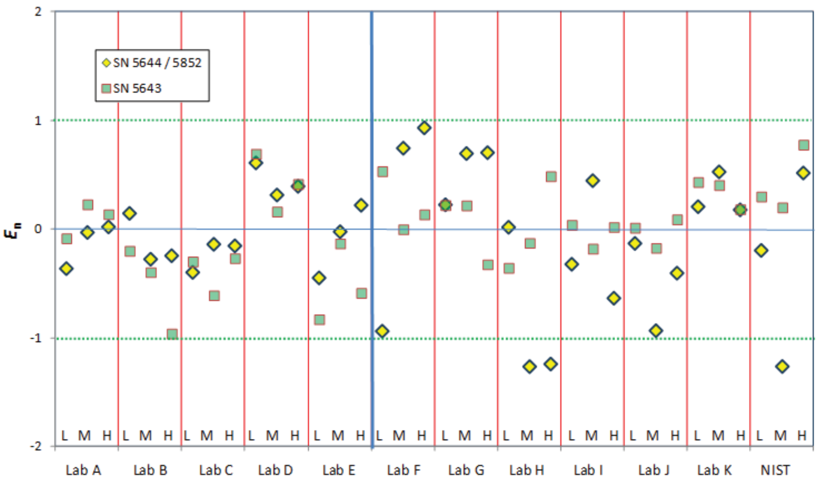


Figure 4. Standardized degree of equivalence for each meter while in the upstream position at the 3 flow set points (low=L, medium=M, and high=H). The vertical line between Labs E and F indicates the change from turbine SN 5644 to SN 5852.

0.1% to 0.14%. A more stable transfer standard is required to evaluate the uncertainty statements of participants in future studies.

Most of the comparison participants used Stoddard solvent, a surrogate for jet fuel, as the test liquid. Several participants instead chose mixtures of propylene glycol and water with the same kinematic viscosity as jet fuel because it is biologically and environmentally benign. Recent theoretical and experimental studies at NIST [8] conclude that the calibration results are effectively the same for either liquid. The results of this comparison are consistent with that conclusion: there was no significant difference between the labs using PG + W and those using Stoddard solvent. Some concerns remain about the long-term effect of exposing 440c stainless steel turbine meter bearings to water solutions, and NIST is now conducting experiments on this topic. To date, we have found that keeping bearing exposure to water to a minimum and drying the meters after calibration with successive ethanol washes is sufficient to prevent corrosion. Where practical, NIST and other laboratories are currently transitioning from Stoddard solvent to safer water-based solutions.

Acknowledgements: The financial support for this study by the Department of Defense Physical/Mechanical Calibration Coordination Group under Project Number 585 is gratefully acknowledged.

References

[1] Mattingly, G. E., “A Round Robin Flow Measurement Testing Program Using Hydrocarbon Liquids: Results for First Phase Testing,” NISTIR 88-4013, September 1988.

[2] *Guidelines for CIPM Key Comparisons*, Committee International des Poids et Mesures, March 1999, updated October 2003, www.bipm.org.

[3] Cox, M. G., “The Evaluation of Key Comparison Data,” *Metrologia*, 39, pp. 589 to 595, 2002.

[4] Wright, J. D., “Uncertainty of the Critical Venturi Transfer Standard Used in the K6 Gas Flow Key Comparison,” FLOMEKO proceedings, Johannesburg, South Africa, September 2007.

[5] Dopheide, D., Mickan, B., Kramer, R., Hotze, H.-J., Vallet, J.-P., Reader-Harris, M., Shaw, J.-H., and Park, K.-A., “Final Report on the CIPM Key Comparisons for Air and Nitrogen Conducted in November 2004 / June 2005 CCM.FF-5b, Metrologia, 2006, 43, Tech. Suppl., 07008.

[6] Johnson, A. N., Crowley, C. J., and Yeh, T. T., Uncertainty Analysis of NIST’s 20 Liter Hydrocarbon Liquid Flow Standard, FLOMEKO proceedings, Taipei, Taiwan, October 2010.

[7] Ball, J. M., Turbine Flow Meter Calibration Using Nonhazardous, Nonpolluting Fluid Mixtures to Simulate Fuels and Lubricants,” *Cal Lab*, Jul Aug Sep 2011, pp. 32 to 36.

[8] Pope, J. G., Wright, J. D., Johnson, A. N., and Moldover, M. R., “Extended Lee Model for the Turbine Meter and Calibrations with Surrogate Fluids,” submitted for publication in *J. Flow Meas. and Instrum*.

[9] Mattingly, G. E., “Flow Measurement Proficiency Testing,” in *Flow Measurement: Practical Guides to Measurement and Control*, 2nd edition, D. W. Spitzer editor, ISA, 2001.

This article was submitted by John D. Wright of NIST, Gaithersburg, MD, on behalf of NIST and contributing authors. Inquiries regarding this study may be directed to john.wright@nist.gov.