Flomeko 2005
6 – 9 June 2005

Results of the North American Natural Gas Flow Calibration Laboratory Comparison: CEESI – SwRI – TCC

Pedro I. Espina, National Institute of Standards and Technology

1 ABSTRACT

The performance of the three commercial laboratories that provide flow meter calibrations to the natural gas industry in North America was evaluated. A tandem flow meter transfer standard package was used and it consisted of commercial flow meters (a turbine and a multi-path ultrasonic) separated by a commercial flow conditioner. Results indicate that the transfer standard package has day-to-day reproducibilities in the 0.37 % level, at a 95 % confidence interval. This level of performance was less than desired given that the calibration laboratories claim uncertainties better than 0.3 %. The results show a consistency in performance among laboratories at the 0.3% to 0.4% level, which indicates that the three calibration laboratories have equivalent performances within the scope of their claimed uncertainties. In addition, the results did not pointed out any significant installation effect at any of the laboratories. The results from one of the laboratory capable of variable pipeline pressure were inconclusive in regards to this parameter influence on flow meter performance.

2 INTRODUCTION

The North American natural gas industry requested the development of a round robin test program to assess the level of equivalency of measurements in the three natural gas flow calibration facilities in North America: the Ventura, Iowa, facilities of the Colorado Engineering Experiment Station, Inc. (CEESI), the Winnipeg, Canada, facilities of TransCanada Calibrations Ltd. (TCC), and the Southwest Research Institute Metering Research Facility (SwRI-MRF) located in San Antonio, Texas. Under the shared sponsorship of the Gas Research Institute (GRI), the National Institute of Standards and Technology (NIST), Measurement Canada, Daniel Industries, and the three calibration facilities, the project sought to establish the degree of equivalency between calibrations provided to the North American natural gas industry. Wilsack & Associates Inc. coordinated this project; the data analysis and this report are provided by NIST.

This paper speaks as to the limitations of this research, conceals the identity the laboratories providing the data, talks about the pressure effects previously seen at the SwRI-MRF, and its conclusions, although not exhaustive, are inclusive.

3 TRANSFER STANDARD PACKAGE DESCRIPTION

The transfer standard package used in this laboratory comparison (see Fig. 1) was composed of a Daniel turbine meter (model 12-20-1-5) in conjunction with a Daniel multi-path ultrasonic flow meter (model 3400-3700-451-2-11938). The package was 300 mm (12”) in diameter, 9.7 m (32’) long. Its flow range was 726 to 6,169 actual m^3/hr (25,401 to 215,911 actual ft^3/hr), which was bounded by the throughput range of the turbine meter. Located between the two flow meters was a CPA 50E flow conditioner.*

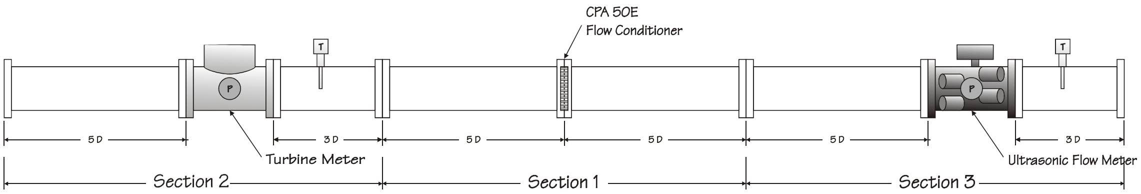

Fig. 1 – Schematic of transfer standard package (flow from left to right).

---

* Certain commercial equipment and materials are identified in this publication to specify adequately the experimental procedure. Such identification does not imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the materials or equipment are necessarily the best available for this purpose.

For operational and transportation requirements, the transfer standard package was divided in three pre-assembled and sealed sections. *Section 1* housed the CPA flow conditioner between two 1.5 m (5') long pipe sections. *Section 2* had a 1.5 m (5') long pipe section prior to the turbine meter, followed by a 0.9 m (3') long pipe section. *Section 3* had a 1.5 m (5') long pipe section prior to the ultrasonic meter, followed by a 0.9 m (3') long pipe section. Pressure was sensed at ports provided in the flow meter bodies, and temperature was sensed at ports located 30 cm (12") downstream of the flow meter gaskets, in their downstream section of pipe.

The transfer standard could be operated in two configurations: assembly A and B. In *assembly A* the ultrasonic flow meter was placed upstream of the CPA flow conditioner, which was followed by the turbine flow meter. For *assembly B* the turbine meter was placed ahead of the CPA, which was followed by the ultrasonic flow meter. The dual configuration of the transfer standard enabled the collection of data that could be used for determining adverse installation effects in the testing facility (given that one flow meter was always exposed to the unconditioned inlet flow of the facility while the other was shielded by the flow conditioner). Further, the presence of two flow meters at all times guarded against possible malfunction of the transfer standard package.

The transfer standard package was instrumented using a sensor package which included two pressure transducers, two frequency counters, and four thermistors. All instruments were connected to the computer using IEEE-488 or IEEE-232, and the data-acquisition and control program was written using National Instruments LabVIEW. The LabVIEW software was designed to sequentially scan all instruments during a 5 seconds interval.

4 RESULTS

The transfer standard package was pre-tested at CEESI Iowa on August 16-22, 2002. The protocol used was designed to answer the following questions:

1. What is the magnitude of the day-to-day reproducibility[†] of the transfer standard package? [1]
2. What is the magnitude of the morning-to-afternoon reproducibility of the transfer standard package?
3. What is the effect of the transfer standard package configuration (i.e., assembly A versus assembly B) in the performance of the transfer standard package?
4. How do installation effects (e.g., stream wise flow swirl) affect the performance of the transfer standard package?
5. What is the magnitude of the transfer standard package hysteresis?[‡]

In this report, only question 1 is considered in detail. However the pre-test data set is available to answer the other questions.

During this phase of the testing program, we were only concerned with the performance of the transfer standard package and thus, we did not collect results from the testing laboratory; a decision precluded the testing laboratory from gaining any advantage from the pre-testing exercise. Table 1 shows the protocol used for the pre-test. Its first column designates the day of testing. As shown, the pre-test lasted 7 working days, which were subdivided into (1) morning and (2) afternoon sets of experiments (see the second column). The third column indicates the transfer standard package configuration during a particular test: assembly A or assembly B (see Fig. 1). The fourth column indicates the presence of an adverse flow meter installation induced by a swirling spool. The particulars of the swirling spool were arbitrarily selected and thus are not important to this discussion; suffice it to say that the spool contained two spiral plates which induced axisymmetrical swirl in the test flow (see Fig. 2). The fifth column indicates the order in that the sequence of test flows was conducted (i.e., ramping the flow up from low to high, or ramping it down), and the sixth column shows the nominal values of the flows tested. These nominal flows had average stream wise velocities of: L ≈ 10 ft/s, M ≈ 30 ft/s, and H ≈ 55 ft/s.

[†] Closeness of the agreement between the results of measurements of the same measurand carried out under changed conditions of measurement [1].

[‡] Level of reproducibility that is achieved when the measurand is approached from a state of larger value versus that is obtained when the measurand is approached from a state of smaller value.

The raw data used in this report was collected as follows. First the transfer standard package was configured and installed in the test facility as per the requirements of the experiment set selected in Table 1. This configuration would be common to all tests conducted during the morning or afternoon of the selected day test. The installation was accomplished by the staff of the test facility and inspected by a member of Wilsack & Associates. A gas leak test was performed while NIST staff verified the proper operation of their instrumentation. The facility operator set the controls to provide the first flow in the testing sequence, and a period of time was observed (about half-hour) to secure stability in pressure, temperature and flow in the test section.

Table 1 – Pre-test protocol.

Day	Assembly	Meter Position (MP)	Induced Facility Effects (FE)	Flow Pattern (FP)	Flow Rate (FR)
1	1	Assembly A	Swirl	Ramp Down	H
					M
					L
				Ramp Up	L
					M
					H
	2	Assembly B	None	Ramp Up	L
					M
					H
				Ramp Down	H
					M
					L
2	1	Assembly A	None	Ramp Up	L
					M
					H
				Ramp Down	H
					M
					L
	2	Assembly A	None	Ramp Down	H
					M
					L
				Ramp Up	L
					M
					H
3	1	Assembly A	Swirl	Ramp Up	L
					M
					H
				Ramp Down	H
					M
					L
	2	Assembly B	Swirl	Ramp Down	H
					M
					L
				Ramp Up	L
					M
					H
4	1	Assembly A	None	Ramp Up	L
					M
					H
				Ramp Down	H
					M
					L
	2	Assembly B	None	Ramp Up	L
					M
					H
				Ramp Down	H
					M
					L
5	1	Assembly A	None	Ramp Up	L
					M
					H
				Ramp Down	H
					M
					L
	2	Assembly A	None	Ramp Down	H
					M
					L
				Ramp Up	L
					M
					H
6	1	Assembly B	Swirl	Ramp Down	H
					M
					L
				Ramp Up	L
					M
					H
	2	Assembly A	None	Ramp Up	L
					M
					H
				Ramp Down	H
					M
					L
7	1	Assembly A	None	Ramp Down	H
					M
					L
				Ramp Up	L
					M
					H
	2	Assembly A	None	Ramp Down	H
					M
					L
				Ramp Up	L
					M
					H
8	1	Assembly C	None	Ramp Up	L
					M
					H
				Ramp Down	H
					M
					L

Because during the pre-test we were not collecting data from the testing laboratory, the transfer package turbine meter frequency was used as the flow control indicator for the test (i.e., the speedometer). The turbine meter was selected over the ultrasonic flow meter because of its higher sensitivity, which was expected to improve the reproducibility of the experiments. This flow control parameter was used throughout this project and thus, the test set points were obtained after the testing laboratory set its flow to yield the following turbine meter frequencies: $f_{T_L} \approx 47\,\mathrm{Hz}$, $f_{T_M} \approx 137\,\mathrm{Hz}$, and $f_{T_H} \approx 240\,\mathrm{Hz}$.

Fig. 2 – Pictures of swirling spool used during the pre-test. This device was used to introduce an axisymmetric component of swirl (i.e., and arbitrary installation effect) into the flow.

Once temperature and pressure in the flowing test section was observed to be stable, the facility operator gave a voice signal to the NIST staff member commencing the formal acquisition of data. Data was acquired over a period of approximately 7 minutes (the time required to acquire the data for three consecutive conventional sample periods of the test facility) and the test concluded with another voice command from the facility operator to the NIST staffer. The process was repeated at two additional flow settings and then, the test section was brought to a zero-flow condition (pressurized) before the second set of three flow settings was tested.

The process described above constituted either a morning or an afternoon in Table 1. Following its completion, the testing facility staff dismantled the transfer standard package into its three main sections (see Fig. 1) in preparation for the next morning or afternoon testing session. To ensure the inclusion of put-in-take-out reproducibility[‡], the transfer standard package was dismantled even if the next assembly configuration was identical to that just tested.

4.1 Grouping of Results

One way to study the results of the experiments in Table 1 is by grouping them in sets of similar character. For example, if we consider the six parameters prescribed in Table 1, it follows that there are only two identical test days: Day-2 and Day-5. As we shall see, data acquired on those two days can be used to estimate a measure of the day-to-day reproducibility of the transfer standard package. Likewise, if we ignore any one test parameter of the set, there are five possible groupings of experiments that can be obtained (e.g., if we ignore test Day there are three sets of identical experiments: set 1 – Day-2.1 + Day-6.2; set 2 – Day-3.2 + Day-6.1; and set 3 – Day-7.1 + Day-7.2).

A similar sub-grouping process can be performed by ignoring the influence of two or even three parameters in Table 1. The resulting sets of experiments provide additional information on the functionality of any set of parameters on other parameters used in the experiment. In the following sections we will discuss some of the conclusions that can be drawn from such sets.

After each test session, the acquired data was inspected for outliers, which were removed prior to averaging, and the resulting averaged data was condensed into tables similar to that shown in Table 2.

[‡] Reproducibility that results from re-testing a flow meter after it has been removed from the pipeline and reinstalled as before.

Table 2 – Sample of averaged data table for a morning set of tests. Shades of gray indicate similar nominal flow settings. For each shaded section, the first row contains the mean value of the collected data and the second row contains its standard deviation. (Note that the second table would appear as a continuation to the first table on its right hand side.)

NIST_12345
8/22/2002
North American Laboratory Comparison Project

Flag	Time	Frequency (Hz)	Q_T (ACMH)	P_{Hi} (psi)	P_{Lo} (psi)	T_1 (°C)	T_2 (°C)
60	0:07:10	238.56	4183.53093	1073.03367	1062.06648	22.8152333	0.0119
		0.30%	*0.30%*	*0.02%*	*0.03%*	*0.03%*	*0.00%*
60	0:07:00	137.44	2410.21779	1076.24451	1072.66247	22.4939153	0.01181356
		0.74%	*0.74%*	*0.01%*	*0.01%*	*0.00%*	*0.00%*
60	0:07:09	47.11	826.166999	1078.78762	1078.49075	22.4223833	0.01176667
		0.95%	*0.95%*	*0.01%*	*0.01%*	*0.00%*	*0.00%*
60	0:07:08	46.32	812.287525	1078.53138	1078.24845	22.5520167	0.01176667
		1.25%	*1.25%*	*0.01%*	*0.01%*	*0.02%*	*0.00%*
60	0:06:53	137.22	2406.36554	1071.97456	1068.40344	23.0162632	0.01170175
		1.34%	*1.34%*	*0.01%*	*0.02%*	*0.01%*	*0.00%*
60	0:06:43	236.97	4155.73731	1066.11718	1055.36039	23.01075	0.01153571
		0.41%	*0.41%*	*0.01%*	*0.01%*	*0.01%*	*0.00%*

T3 (°C)	T4 (°C)	Va (m/s)	Vb (m/s)	Vc (m/s)	Vd (m/s)	Vm (m/s)
24.0337333	24.1022167	-7.4561138	27.5545071	32.6741669	-1.9704352	20.4647615
0.00%	*0.02%*	*-1.93%*	*1.59%*	*0.92%*	*-9.64%*	*1.13%*
24.171661	24.0915593	-4.2356905	16.0769079	19.2282968	-1.0656488	12.0063311
0.00%	*0.00%*	*-2.54%*	*1.91%*	*1.25%*	*-18.96%*	*0.92%*
24.2273	24.1467	-1.4326775	5.46578683	6.5218589	-0.3322578	4.07221802
0.00%	*0.00%*	*-2.04%*	*1.10%*	*1.07%*	*-7.45%*	*1.02%*
24.2253833	24.1409	-1.4149903	5.37018293	6.42924422	-0.3212045	4.00835997
0.00%	*0.00%*	*-2.20%*	*1.54%*	*1.63%*	*-8.49%*	*1.54%*
24.1559649	24.0199123	-4.2494687	15.9669383	19.2091962	-1.0777903	11.9561209
0.00%	*0.00%*	*-2.74%*	*2.49%*	*1.58%*	*-16.46%*	*1.54%*
24.0056429	23.7903929	-7.3752389	27.4576211	32.5741653	-1.9159685	20.4120892
0.00%	*0.00%*	*-1.78%*	*1.32%*	*0.77%*	*-7.21%*	*0.92%*

Ca (m/s)	Cb (m/s)	Cc (m/s)	Cd (m/s)	Cm (m/s)	Q (ACMH)
426.218562	429.533541	429.443725	426.407133	427.900741	5319.59577
0.02%	*0.06%*	*0.04%*	*0.02%*	*0.02%*	*1.13%*
426.920092	428.130133	428.255831	427.059054	427.591277	3120.91733
0.01%	*0.02%*	*0.02%*	*0.01%*	*0.01%*	*0.92%*
427.190204	427.409479	427.597316	427.309222	427.376556	1058.5295
0.00%	*0.01%*	*0.01%*	*0.00%*	*0.00%*	*1.02%*
427.180212	427.399233	427.599757	427.298642	427.36946	1041.93029
0.01%	*0.01%*	*0.01%*	*0.01%*	*0.01%*	*1.54%*
426.747449	427.942853	428.117947	426.88163	427.422471	3107.8657
0.01%	*0.02%*	*0.02%*	*0.01%*	*0.01%*	*1.54%*
426.118109	429.335015	429.33616	426.309807	427.774773	5305.90418
0.02%	*0.06%*	*0.03%*	*0.02%*	*0.02%*	*0.92%*

4.2 Youden Analysis

In this report we make use of *Youden plots* to evaluate results. Before we proceed with a full description of results, it is appropriate to briefly review this statistical analysis method.

If a measurand[♠] is measured using two instruments, one can draw conclusions as to the origins of the error in the measurement process. Taking advantage of such redundant measurement

---

[♠] Particular quantity subject to measurement [1].

process, statistician W. J. Youden developed a graphical analysis tool capable of separating the measurement error into its random[♣] and systematic[♥] contributions [2, 3]. The method is implemented as follows.

In an x-y plot, multiple readings of instrument #1 are normalized by their average and plotted along the x-axis, while multiple readings of instrument #2 are normalized by their average and plotted along the y-axis. Under ideal measurement performance (i.e., no error contributions in the observations), all results should coalesce at $x = 1$ and $y = 1$. Deviations from this behavior along the SW-NE direction indicate that, for those particular tests, both instruments read either higher (i.e., high-high) or lower (i.e., low-low) than their respective averages suggesting that systematic errors affect the measurement process. Deviations along the NW-SE direction indicate that one of the instruments read higher than its average while the other read lower (i.e., high-low or low-high); this type of result suggests that random errors influenced the measurement process.

Typically in a Youden plot, results appear in one of two configurations: a circular pattern originating at $\{1,1\}$ which represents a completely random measurement process, or an oval pattern originating at $\{1,1\}$ and with its principal axis aligned in the SW-NE direction which represents a measurement process exhibiting both random and systematic errors. In the circular pattern, the diameter of the circle is a measure of the randomness of the process. In the oval pattern, the length of the principal axis is a measure of the random plus systematic errors while the length of the secondary axis is a measure of only the random error in the measurement process.

Flow metrology presents difficulties to the implementation of Youden's method because typically two flow meters cannot be physically placed at adjacent locations in the same pipeline, at the same time, without affecting the performance of each other. Mattingly et al. solved this problem by extending Youden's method to measurements made using a tandem flow meter transfer standard like the one here used (see Fig. 1) [4, 5]. Using the tandem transfer standard configuration, one does not compare two instrument readings of the same measurement process but rather, readings of two instruments sensing an identical measurement processes at two different times. For Mattingly's method to be valid, the two measurement processes need to be "identical" to each other within their put-in-take-out reproducibility. In its simplest form, the Mattingly's method could be implemented by replacing flow meter #1 with flow meter #2 and repeating the measurement, however one risks not detecting possible damage to either flow meter between tests (something that could happen in an inter-laboratory comparison). Therefore, Mattingly et al. suggested the tandem arrangement in which both meters are tested at the same time and then swapped for the second set of measurements. A flow conditioner is placed between the flow meters to minimize effects of each others readings.

Mattingly's method produces two Youden plots: one for the flow meters in the upstream position, and a second for the flow meters in the downstream position. The upstream meter readings are affected by pipeline installation effects while the downstream meter readings are shielded from them by the flow conditioner. In either configuration, the ratio of the output from the two flow meters should remain constant provided the sensors remain in good working order.

4.3 Day-to-Day Effects

As seen in Table 1, the pre-test protocol had two identical days of testing: Day-2 and Day-5[†] and the averaged results obtained are shown in Table 3. These results can be used to evaluate day-to-day properties of the transfer standard package.

[♣] Result of a measurement minus the mean that would result from an infinite number of measurements of the same measurand carried out under repeatability conditions, also termed type A uncertainty in current practice [1].

[♥] Mean that would result from an infinite number of measurements of the same measurand carried out under repeatability conditions minus a true value of the measurand, also termed type B uncertainty in current practice [1].

[†] Day 2 of the pre-test protocol was carried out on 8/16/2002; Day 5 was carried out on 8/20/2002.

The first six columns in Table 3 designate the particulars of a test point. From left to right the columns specify: (1) test day, 2 or 5, (2) morning=1 or afternoon=2, (3) transfer standard package assembly A or B, (4) flow installation effects, Y or N, (5) flow ramping progression, up or down, and (6) point in the test sequence, 1–6. The next two columns are: the duration of the test and the average turbine meter frequency. Columns ninth and tenth are the volumetric flows in the turbine and ultrasonic flow meters, respectively. The eleventh column in contains the outcome of the point rejection criteria (described below) and the last two columns show the volumetric flow meter factors, $Q_T / Q_{T_{ref}}$ and $Q_U / Q_{U_{ref}}$, where:

Table 3 – Summary of averaged results for the tests used to determine the day-to-day reproducibility of the transfer standard package. Black cells indicate testing conditions; gray cells indicate measured data; white cells indicate computed results; rows labeled “out” indicate rejected test points; other empty rows are those test points not considered because they were paired with the rejected points.

Day	Set	MP	FE	FP	#	Time (s)	f_T (Hz)	Q_T (ACMH)	Q_U (ACMH)	out	Q_T/Q_{Tave}	Q_U/Q_{Uave}
2	1	A	N	u	1	0:07:01	47.4	830	836		1.0024	1.0025
2	1	A	N	u	2	0:06:59	137.0	2403	2419		0.9999	1.0002
2	1	A	N	u	3	0:07:08	239.4	4199	4209		0.9995	0.9998
2	1	A	N	d	4	0:07:07	239.2	4196	4207		0.9987	0.9994
2	1	A	N	d	5	0:07:15	136.7	2397	2413		0.9974	0.9975
2	1	A	N	d	6	0:06:55	47.2	828	833		0.9989	0.9995
2	2	A	N	d	1	0:06:51	239.2	4195	4201		0.9986	0.9980
2	2	A	N	d	2	0:06:55	137.7	2415	2430	out		
2	2	A	N	d	3	0:07:02	47.0	825	831		0.9959	0.9968
2	2	A	N	u	4	0:06:14	47.1	825	831			
2	2	A	N	u	5	0:06:52	137.2	2407	2425		1.0014	1.0023
2	2	A	N	u	6	0:06:57	239.8	4204	4214		1.0008	1.0010
5	1	A	N	u	1	0:06:56	47.3	829	836		1.0004	1.0026
5	1	A	N	u	2	0:07:09	136.6	2395	2410		0.9967	0.9964
5	1	A	N	u	3	0:07:07	239.8	4204	4214		1.0008	1.0009
5	1	A	N	d	4	0:07:03	239.3	4197	4207		0.9991	0.9993
5	1	A	N	d	5	0:06:53	136.9	2400	2416		0.9988	0.9989
5	1	A	N	d	6	0:06:53	47.2	828	832		0.9993	0.9978
5	2	A	N	d	1	0:06:55	239.9	4207	4212		1.0013	1.0004
5	2	A	N	d	2	0:06:54	137.0	2403	2418			
5	2	A	N	d	3	0:07:00	47.2	827	832		0.9982	0.9982
5	2	A	N	u	4	0:06:58	47.6	836	839	out		
5	2	A	N	u	5	0:06:50	137.1	2405	2420		1.0007	1.0004
5	2	A	N	u	6	0:06:38	239.8	4206	4215		1.0011	1.0012

Q_T is the volumetric flow indicated by the turbine meter (i.e., the averaged turbine frequency divided by the manufacturer meter factor of 5.813 pulses per actual ft^3),

$Q_{T_{ref}}$ is the average of all the Q_T obtained during both testing days, (i.e., $Q_{T_{ref}} = \left(\overline{Q}_{T_{Day2}} + \overline{Q}_{T_{Day5}}\right)/2$)[♦]

[♦] Conventionally the volumetric meter factor is defined as Q_{MUT} / Q_{Lab}, where Q_{MUT} is the reading obtained from the meter under test and Q_{Lab} is the reading obtained from the testing laboratory. However, during the pre-test no results from the calibration facility were recorded which necessitated a definition of the meter factor based on averaged results. For the pre-test

Q_U is the volumetric flow indicated by the ultrasonic flow meter (i.e., obtained directly from the ultrasonic flow meter computer),

$Q_{U_{ref}}$ is the average of all the Q_U obtained during both testing days (i.e., $Q_{U_{ref}} = \left(\overline{Q}_{U_{Day2}} + \overline{Q}_{U_{Day5}}\right)/2$).

As stated earlier, Mattingly's method assumes that the pipeline flow conditions remain unchanged between compared measurements. Thus, at each flow rate, the flow meter factors should not deviate excessively from their mean value. To satisfy this condition a point rejection criteria was applied to the averaged data as follows [6, 7].

At each nominal flow rate, we computed the mean turbine meter frequency, for example using the data in Table 3,

$$\begin{aligned}
\overline{f_T} &= \frac{1}{n}\sum_{i-1}^{n} f_{T\,i} \\
\overline{f_{T_{Low}}} &= \frac{1}{n}\sum_{i-1}^{n} f_{T_{Low}i} = \frac{1}{6}\left(f_{211} + f_{216} + f_{223} + f_{511} + f_{516} + f_{523}\right) \\
\overline{f_{T_{Med}}} &= \frac{1}{6}\left(f_{212} + f_{215} + f_{225} + f_{512} + f_{515} + f_{525}\right) \\
\overline{f_{T_{High}}} &= \frac{1}{8}\left(f_{213} + f_{214} + f_{221} + f_{226} + f_{513} + f_{514} + f_{521} + f_{526}\right)
\end{aligned} \tag{1}$$

where, the three subscript numbers represent the numbers in columns 1, 2, and 6 of Table 3, respectively; we also computed the standard deviation of those same frequencies, $\sigma_{T_{Low}}$.

$$\sigma_{T_{Low}} = \left[\frac{1}{n-1}\sum_{i=1}^{n}\left(f_{T_{Low}i} - \overline{f_{T_{Low}}}\right)^2\right]^{1/2} \tag{2}$$

We then assumed that the dispersion of the frequencies is described by the Gaussian error distribution. This distribution could then be used to compute the probability that any given frequency would deviate a certain amount from the mean frequency. Normally, we do not expect the probability of the deviation to be smaller than $1/2n$ given that this is unlikely to occur in a set of n measurements. Thus if

$$P\left[\left(f_{T_{Low}} - \overline{f_{T_{Low}}}\right)/\sigma_{T_{Low}}\right] < \frac{1}{4n} \tag{3}$$

where P was the probability, then the test point was rejected as an outlier. In simpler terms, if

$$\left|f_T - \overline{f_T}\right| / \sigma_T > 1.327\, n^{0.149} \tag{4}$$

the test point was rejected. In Table 3 we see that points 2-2-2 and 5-2-4 failed the rejection criteria. And because the points are analyzed in pairs, points 5-2-2 and 2-2-4 are also ignored.

Fig. 3 shows the results of the test, with values for Day-2 in the x-axis and values for Day-5 in the y-axis. The square symbols represent results obtained from the upstream meter (i.e., the ultrasonic flow meter), and the circles were obtained from the downstream meter (i.e., the turbine

---

section of this report we used the average results of the day-to-day test, $Q_{T_{ref}}$ and $Q_{U_{ref}}$, as reference values for all results presented. Those values are: $Q_{T_{ref}} = 828$ acmh (for low flow), 2403 acmh (for medium flows), and 4201 acmh (for high flows); $Q_{U_{ref}} = 834$ acmh (for low flow), 2419 acmh (for medium flows), and 4210 acmh (for high flows). Note that for all cases, $Q_{U_{ref}} > Q_{T_{ref}}$ which is unphysical given that the ultrasonic flow meter was in the upstream position in these tests. However, we used the manufacturer meter factors which will be shown later in this report to not be correct.

meter). In the left-hand-side portion of Fig. 3, there are twelve points of each type (for a total of 24 test points) but only ten points of each type were analyzed after rejecting two pairs. On the right-hand-side portion of the figure, we have drawn ellipsoids representing the 95 % confidence level interval (i.e., 2σ) of the considered results about their respective averages.

The results in Table 3 do not fully satisfy all elements of the Mattingly's analysis given that the flow meters were not swapped (notice that all points were acquired using assembly A). Thus, the analysis below considers readings from a single instrument sensing identical measurement processes at two different times. This provides an extra degree of correlation between the day-to-day resulting in a higher contribution to the test measurement error from the transfer standard package (see below).

In the top plots we see that the position of circles and squares appears highly correlated (i.e., the points appear in pairs) because the upstream and downstream results were obtained simultaneously and thus, are correlated in time. There is little distinction between the results obtained in either day with the points arranged in elliptical patterns about the $\{1,1\}$ point. The ellipses are not quite centered at the $\{1,1\}$ point because the individual points were normalized using the average value for both days (i.e., $Q_{T_{ref}}$ or $Q_{U_{ref}}$), while the centers of the ellipses represent the averages for each day for each meter position, hence the slight off-set (see Table 4 for quantitative details).

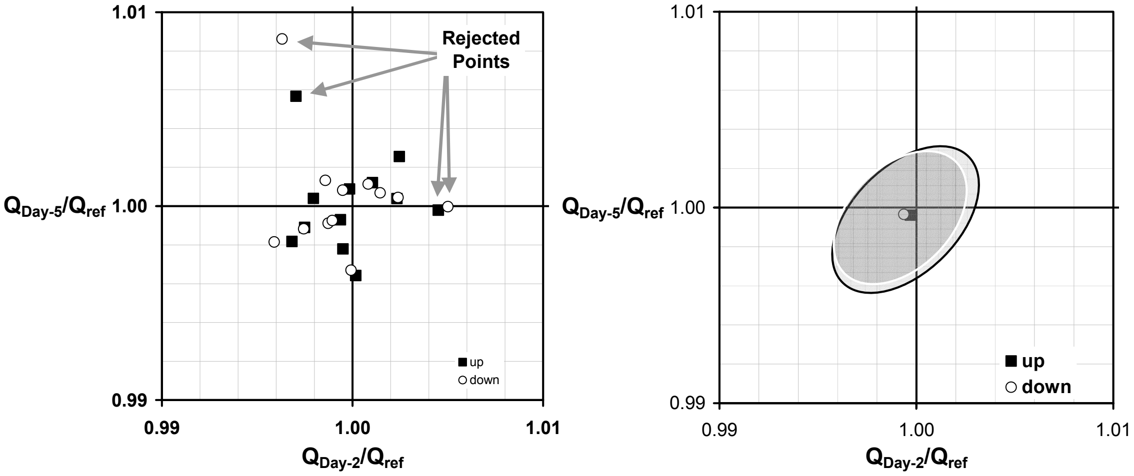

Fig. 3 – Evaluation of day-to-day effects based on Youden's analysis of two identical days of results obtained during the pre-test. *Left* – square symbols are results obtained from the flow meters in the upstream position; circles were obtained from the flow meters in the downstream position. *Right* – ellipsoids represent the 2σ limit of the results after rejecting outliers.

The dispersion of the data is mostly due to random effects in the test as the principal axes of the ellipses are of similar magnitude to their second principal axes (about $1.6:1$). According to Mattingly, the variance of these tests can be separated along that portion contributed by the transfer standard package and that portion contributed by the calibration system. To accomplish this separation, it is necessary to compute the correlation coefficient between the Day-2 and Day-5 results, r_{25}.

The variance decomposition shows that about 53% of the variance $\left(\left|r_{25}\right|_{up}+\left|r_{25}\right|_{down}\right)/2$ is due to the transfer standard package, while 47% $\left(\left[1-\left|r_{25}\right|_{up}\right]+\left[1-\left|r_{25}\right|_{down}\right]\right)/2$ is due to the test facility (see [8] for details on the decomposition process). Based on these results, both meters are capable of reproducing results from day-to-day at $\pm 0.37\%$ of reading, calculated using the relations $\left(\sqrt{\left(\text{stdev}_{2m}\right)_{up}^{2}+\left(\text{stdev}_{5m}\right)_{up}^{2}}\right)$, for the ultrasonic meter and $\left(\sqrt{\left(\text{stdev}_{2m}\right)_{down}^{2}+\left(\text{stdev}_{5m}\right)_{down}^{2}}\right)$ for the turbine meter.

$$r_{25} = \frac{\sum_{i=1}^{n}\left(f_{2i} - \overline{f_2}\right)\left(f_{5i} - \overline{f_5}\right)}{\left[\left\{\sum_{i=1}^{n}\left(f_{2i} - \overline{f_2}\right)^2\right\}\left\{\sum_{i=1}^{n}\left(f_{5i} - \overline{f_5}\right)^2\right\}\right]^{1/2}} \tag{5}$$

Table 4 – Statistics from day-to-day test.

	downstream	upstream			downstream	upstream	
average$_2$	0.9994	0.9997		$\lvert r_{25}\rvert$	0.43	0.50	*system*
average$_5$	0.9996	0.9996		$1-\lvert r_{25}\rvert$	0.57	0.50	*meter*
distance$_{25}$	0.0003						
				stdev$_{2s}$	0.25%	0.27%	2σ
stdev$_2$	0.19%	0.19%		stdev$_{2m}$	0.36%	0.36%	2σ
stdev$_5$	0.15%	0.18%		stdev$_{5s}$	0.10%	0.13%	2σ
variance$_2$	3.65E-06	3.66E-06		stdev$_{5m}$	0.28%	0.34%	2σ
variance$_5$	2.24E-06	3.36E-06					

The day-to-day reproducibility of the results for the transfer package is somewhat larger than the uncertainty claimed by the participating laboratories of this test $(1.23:1)$. Although the ideal laboratory comparison would have transfer standard performance better than that of the laboratories to be tested (~¼ of the uncertainty claimed by the laboratories), the results given below are based on the reproducibilities estimated here.

Laboratory Comparison

With pretest results in hand, we conducted a test protocol in all laboratories to compare laboratory performance under similar volumetric flow conditions.[¥] The 2-day protocol (shown in Table 5) selected for these tests was designed to allow each calibration laboratory to test the individual flow meters in the transfer package using their typical calibration procedure[*] in addition to testing the entire transfer package in its two configurations (i.e., assembly A and B). The normal calibration procedure provided the laboratories with an opportunity to verify the performance of the transfer standard flow meters, whereas the results from tests involving the entire transfer package and agreed-upon test protocol were to provide quantitative information addressing the questions:

(a) what is the level of agreement in the performance of the participating calibration laboratories, and
(b) are differences clearly attributable to the laboratory or to the transfer standard?

Table 5. Test Protocol.

Day	Sub-Assembly	Meter Position (MP)	Flow Pattern (FP)	Flow Rate (FR)
1	1	Flow Conditioner + Turbine	Laboratory's Typical Calibration Procedure	
	2	Assembly A	Ramp Up	L
				M
				H
			Ramp Up	L
				M
				H
2	1	Flow Conditioner + Ultrasonic	Laboratory's Typical Calibration Procedure	
	2	Assembly B	Ramp Up	L
				M
				H
			Ramp Up	L
				M
				H

Prior to the initiation of testing, the calibration laboratories requested that the source of results from this comparison be maintained confidential, therefore, no raw data is provided in this report and individual laboratory results are not identified. However, the raw data and results for each calibration laboratory were reviewed and agreed upon by its staff prior inclusion in this report. Because SwRI could only test the artifact at low and medium flow settings, results are presented in terms of

---

[¥] The pipeline pressure was not controlled at TCC and CEESI, but was maintained at a nominal value of 1000 psi at SwRI. The pipeline temperature was not controlled at any of the calibration facilities.

[*] The procedure conventionally used to calibrate a flow meter. These results are not included in this report

averaged meter factors for all flows tested. This condition requires the assumption that the meter factors remain constant as a function of flow.

The data was collected on the following dates: at CEESI on November 6-7, 2002 and May 14, 2003; at SwRI on October 1-2, 2002 and May 28-29, 2003; and at TCC on August 28-29, 2003. As implied by the above dates, tests were conducted twice at CEESI and SwRI, and only once at TCC due to logistical arrangements. Thus the results from two of the laboratories represent averages of two individual tests while the remaining result represents the average of data from one test only. The results from each laboratory were appropriately weighted to ensure equal contribution to the medians.

Fig. 4 shows the results for the flow meters in the upstream (left) and downstream (right) positions. The results are presented using Youden's analysis where the x-axis is the meter factor for the turbine meter (i.e., the instrument reading divided by the calibration laboratory flow at the turbine meter) and the y-axis is the meter factor for the ultrasonic flow meter. In the figures, the data symbols represent the individual results for each calibration laboratory (at all flow rates). The thick lines are the medians of all results in the comparison (often considered the *best estimate* of the true measurand value [1]) and the thin lines are the average of all the results. The dashed circles about the median and the average, represent the 2σ expansion of root sum square of the day-to-day reproducibility of the transfer standard package plus the claimed uncertainties of the participating laboratories. This quantity can be considered as the expanded uncertainty [9] of the comparison and it is given by,

$$\begin{aligned} U_{Comparison} &= 2\sqrt{\left(R_{TS}/2\right)^2 + \left(U_{CEESI}/2\right)^2 + \left(U_{SwRI}/2\right)^2 + \left(U_{TCC}/2\right)^2} \\ &= 2\sqrt{\left(0.0037/2\right)^2 + \left(0.003/2\right)^2 + \left(0.0025/2\right)^2 + \left(0.003/2\right)^2} \qquad (6) \\ &= 0.62\% \end{aligned}$$

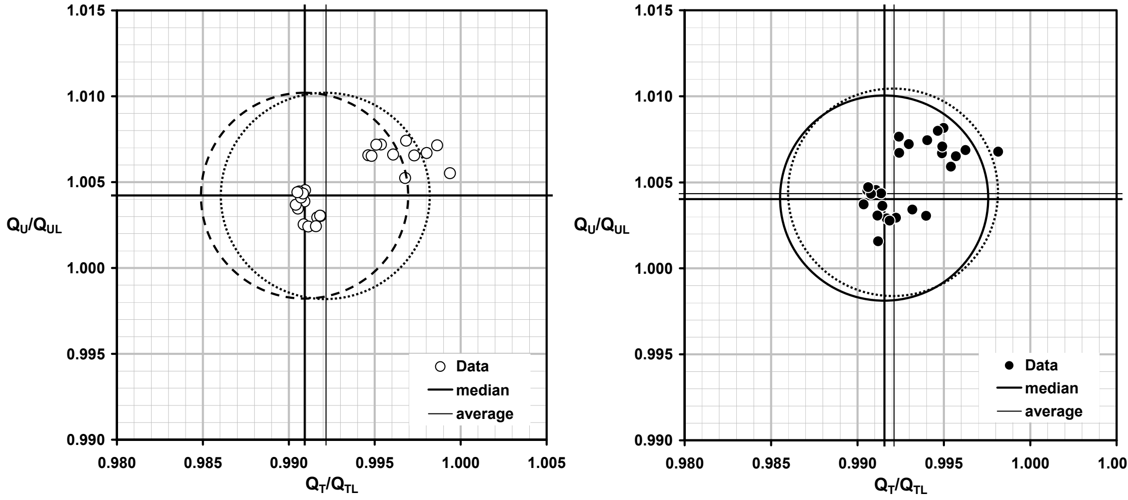

Fig. 4 – Laboratory comparison based on Youden's analysis of results obtained by the flow meters in the upstream and downstream positions. *Left* – upstream results; *right* – downstream results.

Although there are some differences between the upstream meter data (Fig. 4 – *left*) and the downstream meter data (Fig. 4 – *right*), the major features of the up and down stream meters have considerable similarity. Noticeable is a change in the turbine meter response in the results of one of the laboratories (the NE quadrant results). The average response deviation from the median of the turbine was reduced from 0.5% (*left*) to 0.3% (*right*) with the introduction of the flow conditioner, suggesting an installation effect that was reduced by the flow conditioner. However this result is inconclusive as such behavior was not observed for the ultrasonic flow meter.

While there appears to be structure in the data, the source of seemingly persistent differences among the laboratories cannot be determined within the scope of this work. The performance of the transfer package is too poor to clearly distinguish differences in laboratory performance below approximately 0.3%. It is important to remember that the results of any one of the laboratories could prove closer to the true value of the measurand than the results of the other two laboratories. Thus it would be erroneous to conclude that the cluster of results in the NE quadrant implies a calibration error in that facility. More importantly, the results from all laboratories are within the expanded uncertainty of the comparison and thus, we can only conclude that within their claimed uncertainties, these three laboratories provide statistically equivalent results to their customers.

5 PRESSURE EFFECTS

As part of the objectives of this project, the Natural Gas Community expressed interest in assessing the effect pipeline pressure variation on the calibration of a flow meter. This issue is of significance given that typically flow meters are calibrated at a prescribed pipeline pressure, but they are used over a wide range of pipeline pressures in field applications.

The design of the calibration facilities at TCC and CEESI are such that variations of the pipeline pressure are uncontrollable for purposes of calibration where existing pipeline demands sets their pipeline pressure. But as mentioned previously, the design of SwRI's MRF allows for the calibration pipeline pressure to be prescribed. This project took advantage of this capability to compare the performance of the transfer standard package at three line pressures, nominally: 500 psi, 750 psi, and 1000 psi.

The pressure effect tests took place on October 3, 2002 at SwRI. These tests were conducted using the transfer standard package in configuration A (see Fig. 1). Fig. 5 shows the effects of pipeline pressure variation on the volumetric meter factor of the ultrasonic flow meter and the turbine meter. In this figure, the x-axis shows the Reynolds number at each meter (i.e., the Reynolds based on the conditions at each flow meter) and the y-axis shows the volumetric meter factor (i.e., $Q/Q_L = Q_{NIST}/Q_{SwRI}$, where Q_{NIST} is the volumetric flow as measured by the transfer standard). In the figure, error bars represent the 1σ dispersion of those results $(\approx \pm 0.26\%)$.

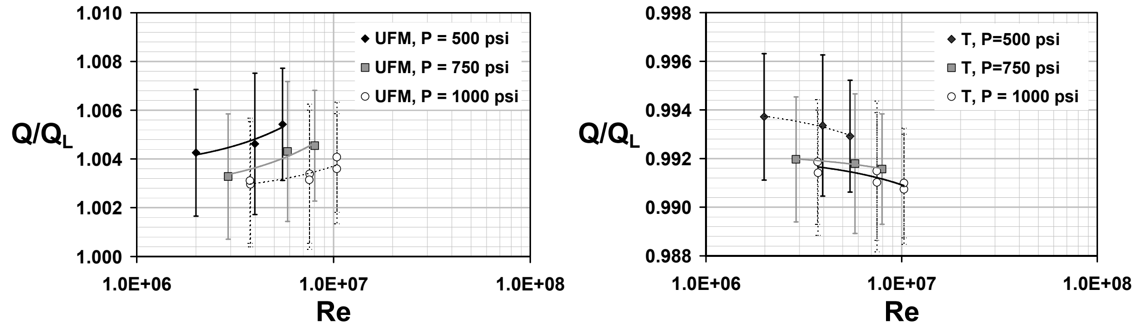

Fig. 5 – Pipeline pressure effect on volumetric meter factor. *Left* – ultrasonic flow meter results; *right* – turbine results.

From Fig. 5 it is apparent that the volumetric meter factor for the ultrasonic flow meter has a numeric value that is consistently larger ($\approx 1.0\%$) than the volumetric meter factor for the turbine. This difference is most likely the result of comparing the data obtained from these meters using their factory default calibration coefficients which are expected to not differ by more than $\pm 0.7\%$ (for the ultrasonic flow meter, [10]) and on the same order for the turbine meter, [11] from their true value. Closer inspection of the data shows that at any given pressure, the volumetric meter factor for the ultrasonic flow meter slightly increases with flow (calibration curves sloping upwards, $\approx 0.1\%$), while, the inverse trend is observed for the volumetric meter factor for the turbine, which slightly decreases with increased flow (calibration curves sloping downwards, $\approx 0.05\%$). However, of more significance is the fact that the calibration curves of both flow meters consistently shift downwards as the pipeline pressure increases.

In Fig. 6 the volumetric flow meter factor of the ultrasonic flow meter as a function of that of the turbine are plotted. Although these plots are strictly not Youden plots, they provide a means to observe similar effects in the results of the MRF tests.

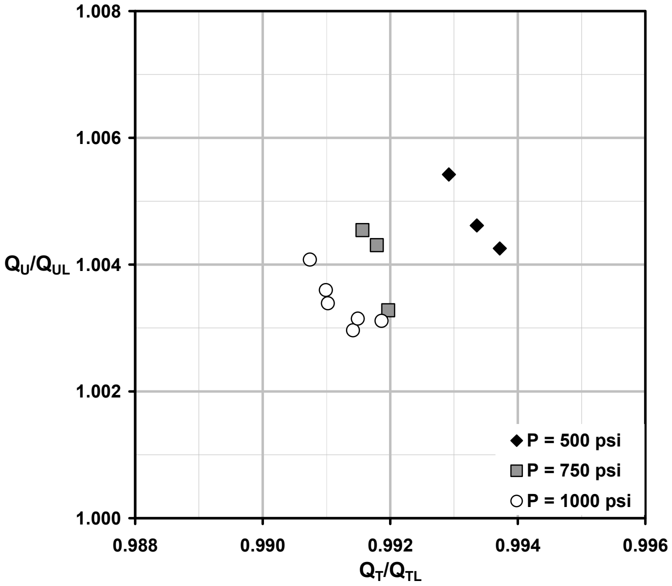

Fig. 6 – Pipeline pressure effect on the volumetric calibration coefficient of flow meters.

From the figure it is clear that the results are clustered by pipeline pressure level and that these clusters move to the NE quadrant of the plot as the pressure decreases. As the pipeline pressure decreases from 1000 psi to 750 psi, the volumetric meter factors of both flow meters increase by $\approx 0.06\%$. Because the flow meters are of different sensing types and yet, their volumetric meter factor shifts are of similar magnitude, suggests that the pressure effect might be induced by the calibration procedures which are the common linkage between the two metering devices. However, when the pipeline pressure is further reduced to 500 psi, the volumetric meter factor for the ultrasonic flow meter shifts by an amount equal to that seen previously, but the volumetric meter factor for the turbine meter shifts more than twice as much, suggesting that each meter responds differently to changes in pressure, and that the pressure effect may be related to the individual meters. Since these results are inconclusive, further investigation of potential sources of these effects, including tests at other laboratories with the ability to vary pressure, appears to be warranted.

6 CONCLUSIONS

A tandem flow meter transfer standard package was used to evaluate the performance of the three commercial laboratories providing flow meter calibrations to the natural gas industry in North America. The transfer standard package made use of a commercial turbine meter and a commercial multi-path ultrasonic flow meter. In the transfer standard package, a commercial flow conditioner separated the flow meters. Results suggest that the transfer standard package has day-to-day reproducibilities in the 0.37% level, at a confidence interval of 2σ. This level was less than desired given that the calibration laboratories participating in the comparison claim uncertainties better than 0.3%.

The laboratory comparison results presented in this report appear to show a consistency in performance among laboratories at the 0.3% to 0.4% level. This is near the limits of detection reasonably expected given the combined performance of the transfer standard and the testing protocols used with it. It should be emphasized that the three calibration laboratories have equivalent performances within the scope of their claimed uncertainties. Finally, results showed that the transfer standard package did not detect any significant installation effect at any of the laboratories.

Results of this study conducted at the only laboratory capable of variable pipeline pressure are inconclusive. Further investigation of potential sources of the observed pressure effects, including tests at other laboratories with the ability to vary pressure, is recommended.

7 ACKNOWLEDGMENT

This investigation was sponsored by the Gas Research Institute, the Chemical Science and Technology Laboratory of NIST, Measurement Canada, Daniel Industries, the Colorado Engineering Experiment Station, Inc., TransCanada Calibrations Ltd., and Southwest Research Institute. Statistical support was provided by the Statistical Engineering Division of NIST. The author thanks the Measurement Technical Committee of the North American Natural Gas Industry for their support. This project could not have been completed without the exceptional dedication of the staffs of SwRI MRF, TCC, and CEESI. The contributions of Donald Ward, William Guthrie, George Mattingly, Aaron Johnson, John Wright, Vikram Gowda, Edmund Xu, and Tsyh-Tyan Yeh from NIST; Dietrich Dopheide from PTB; Larry Fraser and Rene Magnan from Measurement Canada; Martin Schlebach, Kevin Warner, John Lansing, and Klaus Zanker from Daniel; James Witte from El Paso Corporation; Jerry Paul Smith from The Williams Companies, Inc.; Armin Bergmann and Rick Wilsack from Wilsack and Associates; and Chuck French from GRI are gratefully acknowledged.

8 REFERENCES

[1] ISO, "International Vocabulary of Basic and General Terms in Metrology, 2nd Ed.," International Organization of Standardization, Geneva, Switzerland, 1993. (This document is abbreviated VIM.)

[2] W. J. YOUDEN, "Statistical Aspects of the Cement Testing Program," Proceedings of the Am. Soc. Testing Mats., 59 (1959), pp. 1120-1128.

[3] W. J. YOUDEN, "Evaluation of Chemical Analyses on Two Rocks," Technometrics, 1, 4 (1959), pp. 409-417.

[4] G. E. MATTINGLY, P. E. PONTIUS, H. H. ALLION, and E. F. MOORE. "A Laboratory Study of Turbine Meter Uncertainty," NBS SP 484 (Gaithersburg, MD: NBS 1977), pp. 33-54.

[5] G. E. MATTINGLY, W. C. PURSLEY, R. PATON, and E. A. SPENCER. "Steps Toward an Ideal Transfer Standard for Flow Measurement," Flow Measurement of Fluids, eds. H. H. Dijstelbergen and E. A. Spencer, (Amsterdam, The Netherlands: North-Holland Publishing Company, 1978), pp. 543-552.

[6] J. P. HOLMAN and W. J. GAJDA. "Experimental Methods for Engineers, 4th ed." McGraw-Hill Book Company, NY, 1984.

[7] R. E. WALPOLE and R. H. MYERS. "Probability and Statistics for Engineers and Scientists, 3rd ed.," Macmillan Publishing Company, NY, 1985.

[8] G. E. MATTINGLY, "Flow Measurement Proficiency Testing", in *Flow Measurement: Practical Guides for Measurement and Control, 2nd ed.,* D. W. Spitzer, ed., ISA, Research Triangle Park, NC, 2001.

[9] B. N. TAYLOR and C. E. KUYATT. "Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results," NIST Technical Note 1297, National Institute of Standards and Technology (January 1993).

[10] AGA, "Measurement of Gas by Multipath Ultrasonic Meters, TMS Report No. 9," American Gas Association, Arlington, VA, 1998.

[11] AGA, "Measurement of Gas by Turbine Meters, TMS Report No. 7," American Gas Association, Arlington, VA, 1996.