Final report on force key comparison CCM.F-K4.a and CCM.F-K4.b for 4 MN and 2 MN forces

T. W. Bartel
NIST Mass and Force Group

Content submitted to and published by:
Metrologia **49**, *Tech Suppl.* 07003
pp. 1-45 (2012)

U.S. Department of Commerce
John E. Bryson, Secretary

National Institute of Standards and Technology
Patrick D. Gallagher, Director

NIST National Institute of Standards and Technology • U.S. Department of Commerce

DISCLAIMERS

Force Key Comparison CCM.F-K4.a and CCM.F-K4.b
for 4 MN and 2 MN Forces

Final Report

T.W. Bartel
NIST Mass and Force Group
U.S.A.

March 6, 2012

Abstract

This report gives the results for the Comité International des Poids et Mesures (CIPM) key comparisons in force, designated by comparison numbers CCM.F-K4.a and CCM.F-K4.b, for the maximum force values of 4 MN and 2 MN, respectively. Eight National Metrology Institutes, employing nine national force standard machines, participated in these comparisons. The National Institute of Standards and Technology (NIST) served as the pilot institute for these comparisons. Four transducers were employed in two star circulations, with one pair circulated among the participants of CCM.F-K4.a and the other pair circulated for CCM.F-K4.b. The transducers exhibited only minor drift over the circulation time periods; however, the results for one of the transducers circulated for CCM.F-K4.a were limited by excessive sensitivity to orientation about the axis of force application.

Two analysis approaches for calculating Key Comparison Reference Values (KCRV) are presented, one employing a classical consensus mean analysis and the other based on weighted mean calculations. Reasonable correlation in the conclusions for individual transducers is seen between the two approaches. The weighted mean approach is extended to yield a KCRV for each of the two force values from the combined data from all of the associated transducers. One of the seven participants at 4 MN, and two of the nine participants at 2 MN, have expanded uncertainty intervals that lie outside of the expanded uncertainty bands for the corresponding weighted mean KCRVs.

Introduction

Four pair of force metrology key comparisons are being conducted in support of the CIPM Mutual Recognition Arrangement which was initiated in 1999 and has currently been signed by representatives of National Metrology Institutes (NMI) from 45 member states of the Metre Convention. The four sets are: CCM.F-K1.a and CCM.F-K1.b, for the force values of 10 kN and 5 kN, respectively, being piloted by the Center for Metrology and Accreditation (MIKES), of Finland; CCM.F-K2.a and CCM.F-K2.b, for the force values of 100 kN and 50 kN, respectively, being piloted by the National Physical Laboratory (NPL), of the UK; CCM.F-K3.a and CCM.F-K3.b, for the force values of 1.0 MN and 0.5 MN, respectively, being piloted by the Physikalisch-Technische Bundesanstalt (PTB), of Germany; and CCM.F-K4.a and CCM.F-K4.b, for the force values of 4 MN and 2 MN, respectively, being piloted by NIST of the USA.

The protocol for these key comparisons was developed at meetings of the Consultative Committee for Mass and Related Quantities (CCM) Force Working Group held at the Commonwealth Scientific and Industrial Research Organization - National Measurement Laboratory (CSIRO-NML), which is now the National Measurement Institute of Australia (NMIA), in Sydney, Australia in October, 1998 [1] and at the Laboratoire National d'Essais (LNE), Paris, France in May, 1999 [2]. The analysis procedures have been refined at further meetings of the Force Working Group at NIST, Gaithersburg, MD, USA in October, 2001; at the Council for Scientific and Industrial Research (CSIR), Pretoria, South Africa in March, 2004; and at the Centro Nacional de Metrologia (CENAM), Queretaro, Mexico in December, 2007.

Participants

The national metrology institutes that participated in the 4 MN and 2 MN force key comparisons are listed in Table 1, along with the types of force standard machine used and their capacities. One of the originally scheduled institutes for CCM.F-K4.a, CSIR of South Africa, was not able to participate in this comparison because of delays in laboratory refurbishment.

Table 1. List of Participating Institutes

Comparison Identification	Regional Metrology Organization	Participating Institute	Country	Machine Capacity (MN)	Machine Type
CCM.F-K4.a (4 MN range)	Interamerican Metrology System (SIM)	National Institute of Standards and Technology (NIST)	USA	4.45	A
	European Collaboration on Measurement Standards (EUROMET)	Laboratoire National d'Essais (LNE)	France	9	C
		Physikalisch-Technische Bundesanstalt (PTB)	Germany	16.5	B
		National Physical Laboratory (NPL)	UK	5	B
	Asia Pacific Metrology Program (APMP)	National Institute of Metrology (NIM)	China	20	B
		National Metrology Institute of Japan / Advanced Industrial Science and Technology (NMIJ/AIST)	Japan	20	B
		Korea Research Institute of Standards and Science (KRISS)	Korea	10	C
CCM.F-K4.b (2 MN range)	SIM	NIST	USA	4.45	A
	EUROMET	PTB	Germany	2	A
		Glowny Urzad Miar / Central Office of Measures (GUM)	Poland	3	C

Machine Type :
A = deadweights alone
B = deadweights with force multiplication
C = hydraulic actuation with reference transducers

Comparison Protocol

The key comparisons were carried out by circulating a pair of force transfer standard transducers among the participating institutes in a star pattern – such that the transducers were returned to the pilot institute by each participant before being circulated to the next. The same measurement procedure was conducted by each participant, including the pilot institute which repeated the measurements each time the transducers were returned. Transportation of the transducers was accomplished by air freight shipment.

For comparison CCM.F-K4.a, the following two transducers were circulated: (a) a Gassmann Theiss Messtechnik (GTM) GmbH Series KTN-D Force Transfer Standard [1], capacity 5 MN, serial number 1608, and (b) a Höttinger Baldwin Messtechnik (HBM) Type RTN Force Transfer Standard, capacity 4 MN, serial number 2813TO. The output voltage ratios of these two transducers at 4 MN are about 1.6 mV/V and 2.0 mV/V,

---

[1] Identification of the equipment listed does not imply recommendation or endorsement by NIST, nor does it imply that the equipment identified is necessarily the best available for the purpose.

respectively. These two transducers are denoted as Transducer 1 and Transducer 2 in the remainder of this report.

Comparison CCM.F-K4.b was begun after the circulation for the 4 MN comparison was completed. For this 2 MN comparison the following two transducers were circulated: (a) a GTM GmbH Series KTN-D Force Transfer Standard, capacity 2 MN, serial number 46163, and (b) a Revere Transducers Inc (RTI) Model CSP-D3-500K-20MH Force Transfer Standard, capacity 2.22 MN, serial number 438481. The output voltage ratios of these two transducers at 2 MN are about 2.0 mV/V and 1.8 mV/V, respectively. These two transducers are denoted as Transducer 3 and Transducer 4 in the remainder of this report.

In addition to the two transducers circulated to each institute, an HBM Model BN100 Bridge Calibration Unit, serial number 8491 was also circulated in order to obtain comparison calibrations for each of the measuring amplifiers used by the institutes to acquire the transducer responses. All participating institutes employed the same make and model of measuring amplifier, an HBM Model DMP 40. As the bridge calibration unit required an input supply voltage of 120 volt AC at 60 Hz, a frequency converting transformer was included to transform the supply voltage of other countries to the required parameters.

A uniform measurement procedure to be followed for all participants was established by the CCM Force Working Group in order to minimize the effects of transducer characteristics, such as hysteresis, creep, and sensitivity to non-axial loading. The procedure involved an unbroken sequence of loading cycles, with forces of 0 MN, 2 MN, and 4 MN for comparison CCM.F-K4.a and 0 MN and 2 MN for comparison CCM.F-K4.b. The sequence incorporated two repetitions of six orientations of the transducer about its vertical axis. At each orientation, two identical loading cycles were conducted, with an unanalyzed exercise cycle preceding the data cycle used to acquire the transducer readings to be analyzed. The established protocol called for all force points to be spaced at six-minute intervals. This timing was adhered to for all transitions except for the 4 MN to 0 MN transition at the end of each cycle for CCM.F-K4.a; this transition was lengthened to nine minutes because of the unloading time requirements of the NIST deadweight machine.

The actual forces applied by the NIST 4.448 MN deadweight force standard machine, which are adjusted to integer values of kilopound-force (klbf), are 2.0017 MN and 4.0034 MN, corresponding to weights of 450 klbf and 900 klbf, respectively. Thus the NIST forces exceed the nominal comparison forces of 2 MN and 4 MN by 0.085 %. NIST's results for the comparison forces were computed by adjusting NIST's readings for this 0.085 % difference. The adjustments made use of a calibration equation relating force to response derived for each transducer from separate measurements, which provided the slope of the response curve at the 2 MN and 4 MN force points.

The NIST machine applies forces by lifting weights that are linked in a manner so as to load sequentially. Each weight increment increases the force by 222.4111 kN. Nine of these weights are applied over a three minute period to increase force from 0 MN to 2.0017 MN, and nine more are applied over another three minute period to increase force from 2.0017 MN to 4.0034 MN. A 6.3 min period is required to unload all 18 weights to decrease force from 4.0034 MN back to 0 MN. This loading and unloading cycle is diagrammed in Fig. 1. The train of measurement cycles for comparison CCM.F-K4.a, without the individual weight increments depicted, is diagrammed in Fig. 2. The dots on the diagrams of each figure indicate the points where the indicator readings are taken.

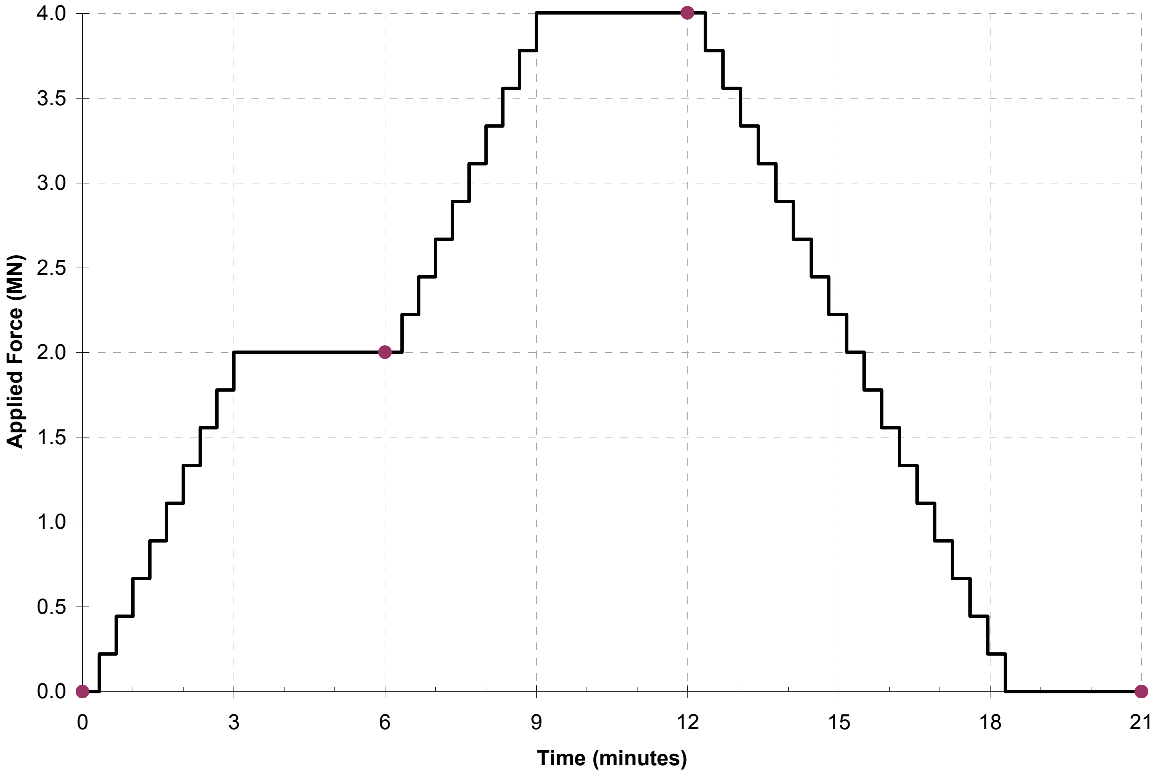

Figure 1. NIST 4.448 MN Deadweight Machine Loading Cycle for CCM.F-K4.a

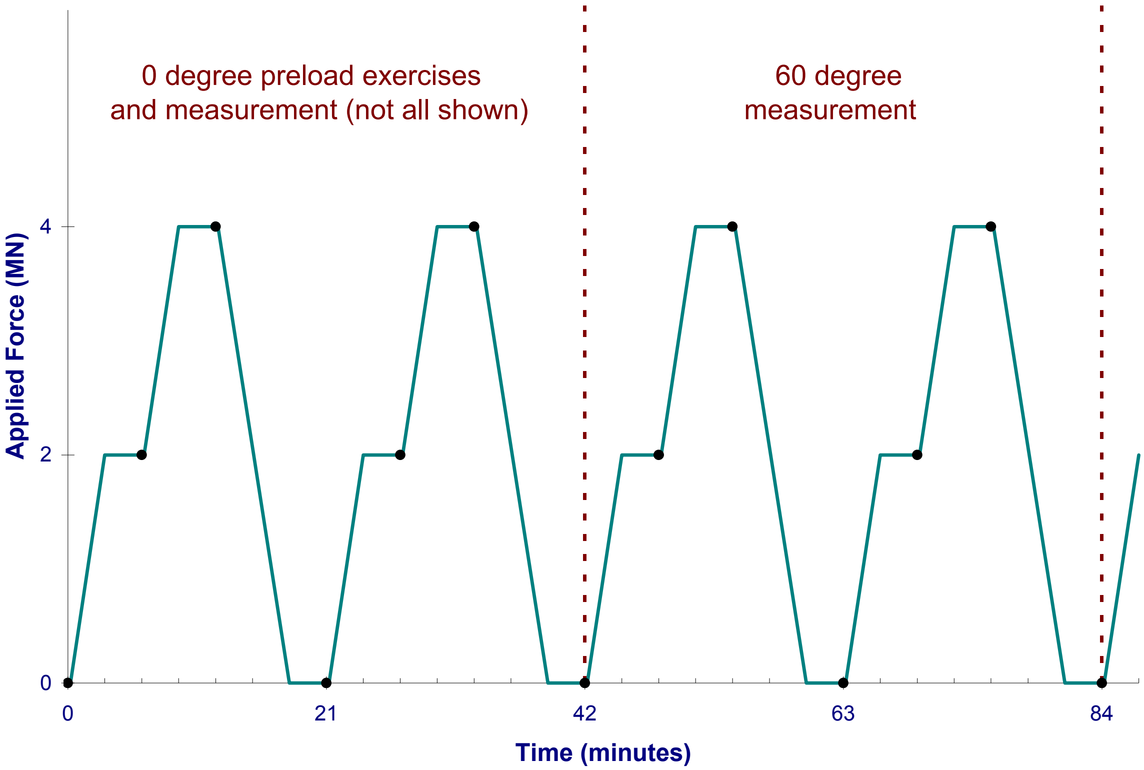

Figure 2. Measurement Sequence for CCM.F-K4.a

The measurements for CCM.F-K4.a and CCM.F-K4.b were conducted with the same deadweight force standard machine. The measurement sequence for CCM.F-K4.b is similar to that shown in Fig. 2, except that the force was returned to zero after the reading at the 2 MN step, and all force points could be spaced at six-minute intervals. The total time for the sequence of several preload cycles at the initial 0º position, followed by two repetitions through the six orientations was 11 h 12 min for CCM.F-K4.a and 6 h 24 min for CCM.F-K4.b.

The "idle" period (almost three minutes) upon returning to zero at the end of the cycle was used to rotate the transducer, for those cycles where the schedule called for such reorientation. Because of the size and weight of the transducers, and the size of the NIST machine, it can require two minutes to accomplish the reorientation and precise realignment of the load cell on the compression platen. Thus the rotation maneuver is finished before the zero-load reading is taken, which has no effect on the analysis because the zero-load reading before each data cycle, rather than after, is used to calculate the deflections.

Stability of the Bridge Calibration Unit

The measurement protocol followed at each participating institute included sampling, by the institute's DMP 40 measuring amplifier, of the BN100 bridge calibration unit that traveled with the force transducers. This sampling, performed at 0.2 mV/V intervals from -0.2 mV/V to 2.2 mV/V using the BN100, was used to correct for differences among the institutes' DMP 40 instruments. This correction procedure relied on the stability of the traveling BN100 over the course of any round trip from NIST to an institute and back.

The stability of the BN100 over the time frame of the entire comparison is depicted in Figs. 3 and 4 for two representative bridge output values of 1.0 mV/V and 2.0 mV/V, respectively. The plots for other bridge output values are similar to the plots shown. These data were obtained with NIST's DMP 40 during the measurement period each time the circulated instruments were returned to NIST. The cluster of four points for each measurement set represents the BN100 readings at the beginning and end of the long measurement day for each of the two force transducers being circulated. In some cases fewer than four points may be visually identifiable in a cluster due to overlapping. The figures give both the recorded readings at the indicated setting, and the net readings obtained by subtracting the corresponding readings at 0.0 mV/V from the recorded readings. The broken lines connect the mean values of the clusters, and indicate the travel time from NIST to another lab and back to NIST. Note that November 2004 presents two closely spaced clusters of four points each, corresponding to NIST's measurements at the end of the circulation for CCM.F-K4.a and the beginning of the circulation for CCM.F-K4.b.

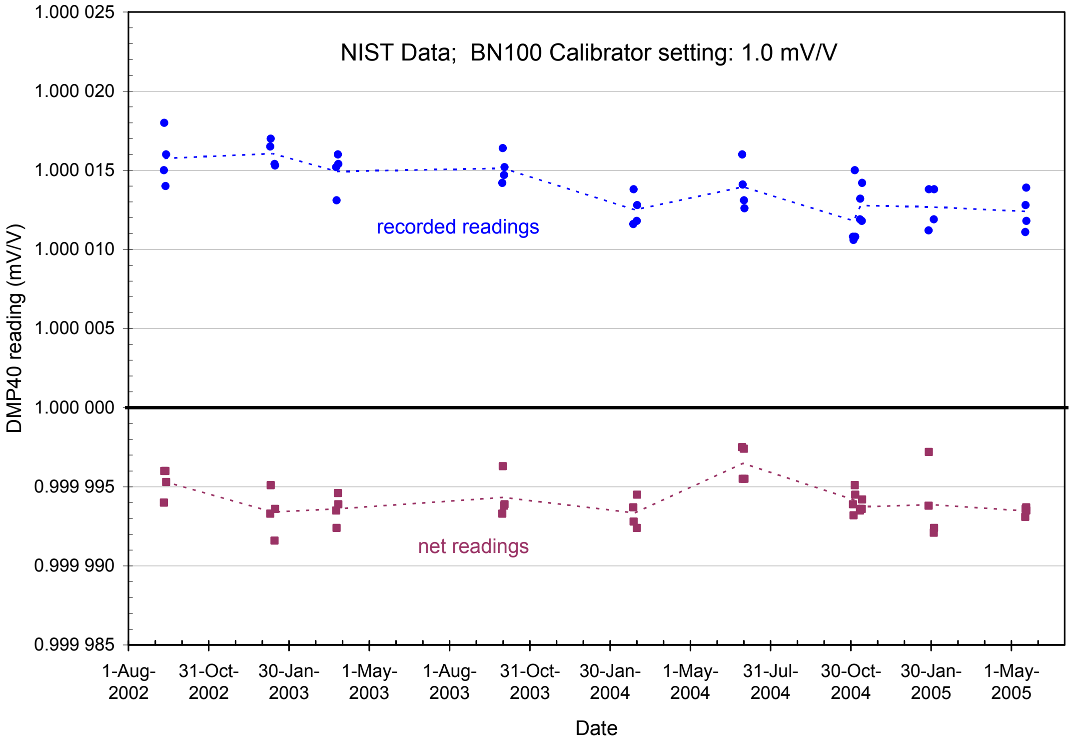

Figure 3. BN100 data obtained at NIST throughout time frame of comparison, for an output value of 1.0 mV/V.

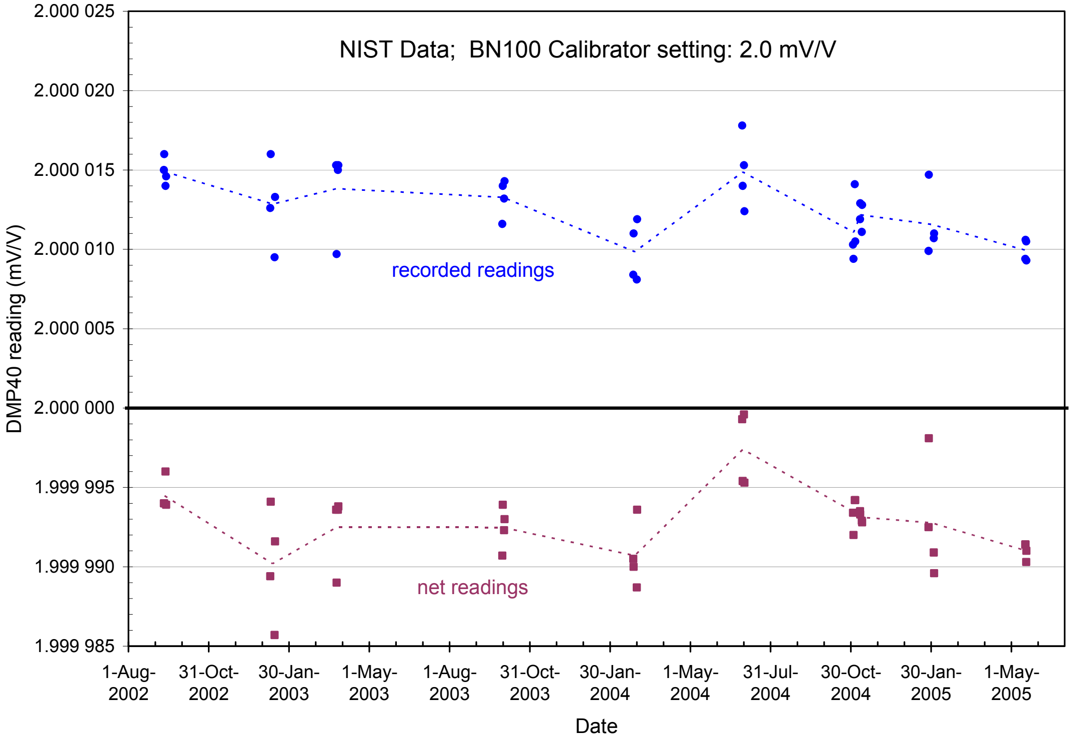

Figure 4. BN100 data obtained at NIST throughout time frame of comparison, for an output value of 2.0 mV/V.

The traveling BN100 is seen to be stable within a relative value of 0.0005 % for any two consecutive measurement sets, representing a round trip to one of the other institutes. Because of the star circulation pattern, only the drift between adjacent measurement sets is relevant. This drift is seen to be of the same order of magnitude as the variation among the readings in the individual measurement sets.

NIST also maintained a second BN100 that remained under controlled laboratory conditions at NIST during this circulation. The readings from this second BN100 were equally stable. The spread in the clusters indicates the limit to which the DMP 40 can be calibrated by the BN100 at any point in time. An estimated standard uncertainty (k=1) in the BN100 readings, relative to the output value, of 0.0005 % is incorporated into the combined standard uncertainty for each laboratory by the symbol u_v in Eq.(4).

Analysis Details

The force transfer standards that were circulated among the participants do not have intrinsically known responses to the applied forces. The analysis was conducted to make use of the devices as comparators, in order to infer a comparison of the participants' force standards at the 2 MN and 4 MN force points for CCM.F-K4.a, and at the 2 MN force point for CCM.F-K4.b. For each of the two transducers employed for each comparison, separate analyses were conducted at each of the two force points.

For each participant, the response r_i was calculated by subtracting the indicator reading at 0 MN at the beginning of each of the twelve data cycles from the indicator readings at 2 MN or 4 MN in the same cycle. The indicator readings incorporate corrections for the offset between each participant's measuring amplifier and the measuring amplifier used at NIST, as determined from data acquired from the bridge calibration unit circulated along with the transducers. The mean response r and standard deviation s for the individual responses r_i , for i = 1 to 12, was computed from

$$r = (1/\mathrm{n})\sum r_i \qquad (1)$$

$$s = [\sum(r_i - r)^2/(\mathrm{n}-1)]^{1/2} \quad , \qquad (2)$$

where n = 12, for each participant at 2 MN and 4 MN for CCM.F-K4.a and at 2 MN for CCM.F-K4.b. The value n = 12 arises from two independent measurements of the transducer response at each of six orientations about its vertical axis. Separate values of r and s are calculated for each of the two transducers circulated to each participant.

A standard uncertainty u_a (k=1), incorporating only the standard deviation s from the comparison measurement data sets, is calculated for each value of r as

$$u_a = [s^2/\mathrm{n}]^{1/2} \ . \qquad (3)$$

This “data-based” standard uncertainty u_a is useful as an indicator of the ability of the transducer, employed as specified by the comparison measurement protocol, to resolve differences in the values of r calculated from Eq.(1) for data sets acquired at different times or by different laboratories. The quantity u_a incorporates uncertainties associated with the characteristics of the transducers, inherent in their design and manufacture, which were selected to serve as force transfer standards for the comparison; u_a also incorporates any uncertainties associated with the implementation of the uniform measurement procedure conducted at each laboratory.

Other sources of uncertainty include: (1) the standard uncertainty in the applied force, denoted by u_f, which incorporates uncertainties associated with the force standard machine employed at each participating laboratory, and (2) the standard uncertainty in determining the measuring amplifier corrections, denoted by u_v, which represents the uncertainty in determining the offsets among the instruments used by each laboratory to acquire the transducer responses r_i.

The values of u_f were obtained from information supplied by the participants from their own uncertainty analyses for their respective force standard machines. NIST has determined the value of u_f for the forces applied by its 4.448 MN deadweight machine to have a relative value of 0.0005 % of the applied force, as described in reference [3]. u_v was estimated to have a value of $(0.000005)r$, for each mean response r, based on repeated measurements conducted with the bridge calibration unit at NIST.

A relevant environmental source of uncertainty is the temperature of the transducer, which has a temperature dependent response. This uncertainty was minimized by the procedural specification that all measurements be conducted at a temperature of 20.0 ºC ± 0.2 ºC. From separate measurements conducted at NIST of the thermal sensitivities of the transducers, and from the temperatures reported by each laboratory during their measurements, it was determined that the uncertainty associated with temperature was not significant.

Thus the three uncertainties u_a, u_f and u_v incorporate all known uncertainties relevant to the comparison which can be quantified. A combined standard uncertainty u_c (k=1), incorporating the standard uncertainties u_a, u_f and u_v, is calculated for each value of r given by Eq.(1) as

$$u_c = [u_a^2 + u_f^2 + u_v^2]^{1/2} \ , \qquad (4)$$

where the three standard uncertainties are expressed in the unit of the response r, which has the unit for the readings returned by the measuring amplifiers, giving the voltage ratio in mV/V.

The entire measurement and analysis procedure was repeated at the pilot institute, NIST, upon return of the transducers from one participant before sending them out to the next.

In order to compensate for any drift in the transducer response over time, the final value to be compared for each participant consisted of the difference between the participant's mean response, r, given by Eq.(1), and the average of the two mean responses for the measurements performed at NIST preceding and following the measurements at the participant's laboratory.

Table 2 through Table 4 give the numerical values for the mean response, r, and standard deviation, s, calculated from Eq.(1) and Eq.(2), respectively, for each measurement data set obtained from the participants for all four transducers. These values are given in the units of voltage ratio. Corrections for measuring amplifier offsets have been incorporated into the responses shown in the tables. Also given in the tables are the values of the standard uncertainty, u_f, in the applied force, provided by the participants for their force standard machines as described in the paragraphs below Eq.(3).

In the remainder of this report, the participating NMIs are not identified by name, but rather are represented by arbitrary lab numbers. These lab numbers denote the participating NMIs as follows:

Lab 1 – NIST
Lab 2 – LNE
Lab 3 – NMIJ/AIST
Lab 4 – KRISS
Lab 5 – PTB (16.5 MN-K-NME machine)
Lab 6 – NIM
Lab 7 – NPL
Lab 8 – GUM
Lab 9 – PTB (2 MN-K-NME machine)

It is noted that Transducer 1 and Transducer 2, appearing in the tables for comparison CCM.F-K4.a, were circulated among NIST and six other participants, whereas Transducer 3 and Transducer 4, used for comparison CCM.F-K4.b, were circulated among NIST and two other participants.

Table 2. Measurement Results for CCM.F-K4.a, Transducer 1

		2 MN force point, Transducer 1			4 MN force point, Transducer 1		
NMI	date of measure-ment set	mean corrected response (mV/V)	data set standard deviation (mV/V)	lab-provided standard uncertainty in applied force (mV/V)	mean corrected response (mV/V)	data set standard deviation (mV/V)	lab-provided standard uncertainty in applied force (mV/V)
Lab 1	9/12/2002	0.799 200	0.000 010	0.000 004	1.598 715	0.000 018	0.000 008
Lab 2	10/25/2002	0.799 215	0.000 016	0.000 200	1.598 764	0.000 020	0.000 400
Lab 1	1/9/2003	0.799 177	0.000 006	0.000 004	1.598 698	0.000 012	0.000 008
Lab 3	1/28/2003	0.799 098	0.000 004	0.000 035	1.598 457	0.000 011	0.000 070
Lab 1	3/25/2003	0.799 190	0.000 014	0.000 004	1.598 720	0.000 034	0.000 008
Lab 4	6/23/2003	0.799 170	0.000 021	0.000 080	1.598 716	0.000 047	0.000 160
Lab 1	9/30/2003	0.799 199	0.000 009	0.000 004	1.598 731	0.000 015	0.000 008
Lab 5	11/27/2003	0.799 161	0.000 013	0.000 028	1.598 672	0.000 018	0.000 056
Lab 1	2/26/2004	0.799 179	0.000 010	0.000 004	1.598 698	0.000 019	0.000 008
Lab 6	5/13/2004	0.799 217	0.000 021	0.000 027	1.598 748	0.000 031	0.000 053
Lab 1	6/29/2004	0.799 192	0.000 009	0.000 004	1.598 720	0.000 016	0.000 008
Lab 7	9/1/2004	0.799 412	0.000 036	0.000 080	1.598 974	0.000 056	0.000 400
Lab 1	11/2/2004	0.799 194	0.000 009	0.000 004	1.598 730	0.000 016	0.000 008

Table 3. Measurement Results for CCM.F-K4.a, Transducer 2

		2 MN force point, Transducer 2			4 MN force point, Transducer 2		
NMI	date of measure-ment set	mean corrected response (mV/V)	data set standard deviation (mV/V)	lab-provided standard uncertainty in applied force (mV/V)	mean corrected response (mV/V)	data set standard deviation (mV/V)	lab-provided standard uncertainty in applied force (mV/V)
Lab 1	9/10/2002	0.999 468	0.000 218	0.000 005	1.999 813	0.000 405	0.000 010
Lab 2	10/24/2002	0.999 450	0.000 151	0.000 250	2.000 013	0.000 260	0.000 500
Lab 1	1/14/2003	0.999 564	0.000 207	0.000 005	2.000 012	0.000 415	0.000 010
Lab 3	1/31/2003	0.999 518	0.000 136	0.000 044	1.999 900	0.000 238	0.000 088
Lab 1	3/27/2003	0.999 556	0.000 149	0.000 005	2.000 005	0.000 329	0.000 010
Lab 4	7/3/2003	0.999 568	0.000 253	0.000 100	2.000 059	0.000 469	0.000 200
Lab 1	10/2/2003	0.999 541	0.000 163	0.000 005	2.000 000	0.000 346	0.000 010
Lab 5	12/9/2003	0.999 493	0.000 169	0.000 035	2.000 008	0.000 274	0.000 070
Lab 1	3/1/2004	0.999 535	0.000 171	0.000 005	1.999 983	0.000 322	0.000 010
Lab 6	5/17/2004	0.999 247	0.000 106	0.000 033	1.999 725	0.000 146	0.000 067
Lab 1	7/1/2004	0.999 540	0.000 154	0.000 005	1.999 987	0.000 307	0.000 010
Lab 7	9/6/2004	0.999 731	0.000 231	0.000 100	1.999 900	0.000 332	0.000 500
Lab 1	11/4/2004	0.999 574	0.000 183	0.000 005	2.000 061	0.000 369	0.000 010

Table 4. Measurement Results for CCM.F-K4.b, Transducer 3 and Transducer 4

		2 MN force point, Transducer 3			2 MN force point, Transducer 4		
NMI	date of measure-ment set	mean corrected response (mV/V)	data set standard deviation (mV/V)	lab-provided standard uncertainty in applied force (mV/V)	mean corrected response (mV/V)	data set standard deviation (mV/V)	lab-provided standard uncertainty in applied force (mV/V)
Lab 1	11/10/2004	1.982 331	0.000 021	0.000 010	1.803 627	0.000 067	0.000 009
Lab 8	12/23/2004	1.981 115	0.000 047	0.000 495	1.802 609	0.000 044	0.000 451
Lab 1	1/27/2005	1.982 312	0.000 028	0.000 010	1.803 634	0.000 068	0.000 009
Lab 9	3/24/2005	1.982 482	0.000 026	0.000 020	1.803 498	0.000 041	0.000 018
Lab 1	5/17/2005	1.982 379	0.000 033	0.000 010	1.803 649	0.000 049	0.000 009

In Tables 2 and 3, the columns labeled "lab-provided standard uncertainty in applied force" indicate considerably lower values for the force standard machine of the pilot laboratory, NIST, than for the force standard machines of the other participants. The reason for this is that the NIST force standard can apply forces to 4.448 MN which are derived from deadweights alone; the force standards of the other participants of CCM.4-K4.a make use of (a) smaller deadweight forces which are multiplied by means of a lever or hydraulic mechanical advantage system, or (b) direct hydraulic actuation to apply forces that are measured with a set of reference transducers. Additional sources of uncertainty apply to these latter two methods of force application.

An additional observation from Tables 2 and 3 is that, of the two transducers employed in CCM.4-K4.a, Transducer 2 yields much higher values for the data-set standard deviation than Transducer 1. Transducer 2 demonstrates a much greater sensitivity to orientation about the vertical loading axis within the force standard machine of each participant. For Transducer 2, this data-set standard deviation, u_a, is the dominant uncertainty component for all but one of the participants, whereas for Transducer 1, the standard uncertainty in the applied force, u_f, is the dominant uncertainty component for all but the pilot laboratory, NIST.

For the results of comparison CCM.F-K4.b, as shown in Table 4, the differences in the data-set standard deviation between Transducers 3 and 4 are less pronounced.

NIST Results

The results for the measurements at NIST alone are shown in Fig. 5 through Fig. 10, in order to show the variation associated with NIST repeatability or transducer drift over time. Figure 5 through Fig. 8 are from comparison CCM.F-K4.a, for Transducer 1 and Transducer 2 with force points at 2 MN and 4 MN. Figure 9 and Fig. 10 are from comparison CCM.F-K4.b, for Transducer 3 and Transducer 4 with one force point at 2 MN. The figures show the differences between the responses for the individual NIST data sets and the mean of the responses for all of the NIST data sets. Note that for ease of comparison of the four transducers, Figures 5 through 10 all have the same vertical scale (-200 to +150, times 10^6 relative to the NIST global mean).

If the index j is used here to indicate only the NIST measurement sets, with $j = 1$ to 7 for CCM.F-K4.a and $j = 1$ to 3 for CCM.F-K4.b, the response differences are

$$d_j = r_j - R_{\mathrm{NIST}} \ , \qquad (5)$$

where r_j represents the j^{th} measurement set mean as computed in Eq.(1) for the NIST data set j, and R_{NIST} is referred to as the NIST global mean:

$$R_{\mathrm{NIST}} = (1/m)\sum r_j \ . \qquad (6)$$

The number of NIST data sets, m, is 7 for CCM.F-K4.a and 3 for CCM.F-K4.b. The ordinates in the plots, D_j, are the response differences relative to the NIST global mean, with a multiplier of 10^6 used to adjust the scale to presentable values (thus $D_j = (10^6)d_j/R_{\mathrm{NIST}}$). The baseline (ordinate 0) for each plot represents the NIST global mean for the data sets on that plot.

Two expanded uncertainty intervals (k=2) are shown for each point in Fig. 5 through Fig. 10. Each left-side (solid line) bar represents the data-based expanded uncertainty $U_{aj} = 2u_{aj}$ for the corresponding data set j, where u_{aj} is calculated from Eq.(3). Each right-side (dashed line) bar represents the total expanded uncertainty $U_{tj} = 2u_{cj}$ for the corresponding data set j, where u_{cj} is calculated from Eq.(4). The uncertainty bars are plotted in the figures as relative to the NIST global mean, and thus have lengths of $(10^6)U_j/R_{\mathrm{NIST}}$.

The left-side data-based uncertainty intervals indicate the sufficiency of the transducer-measurement protocol combination to resolve differences among data sets for a particular transducer. The total uncertainty intervals to the right of each data point indicate the significance of the differences in light of all relevant uncertainty components.

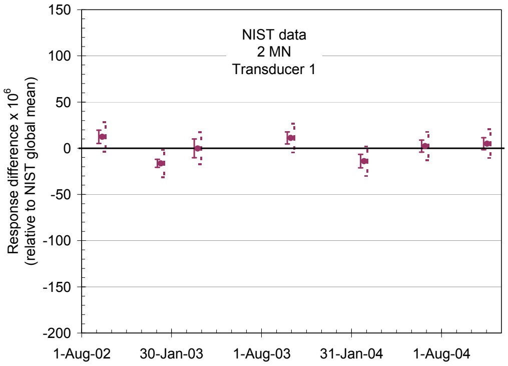

Figure 5. NIST data and *k*=2 expanded uncertainty intervals at 2 MN force point for Transducer 1

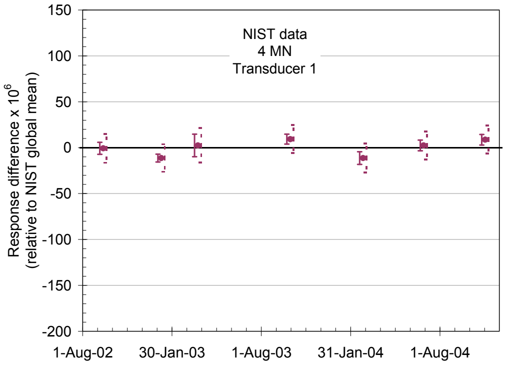

Figure 6. NIST data and *k*=2 expanded uncertainty intervals at 4 MN force point for Transducer 1

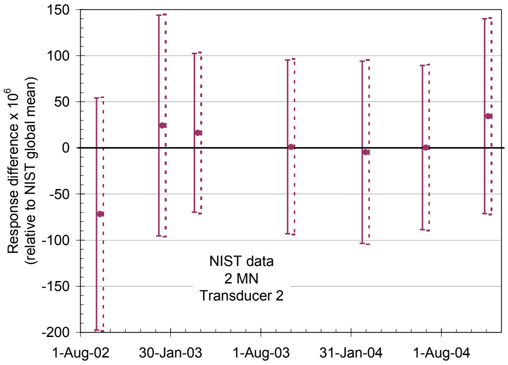

Figure 7. NIST data and *k*=2 expanded uncertainty intervals at 2 MN force point for Transducer 2

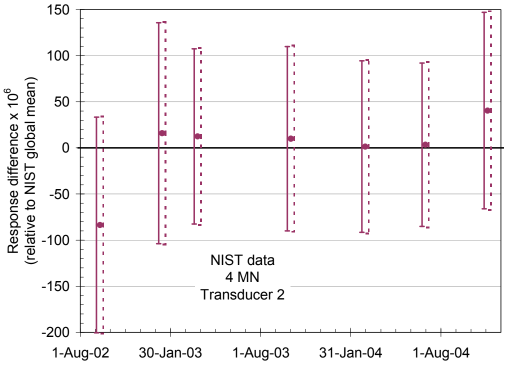

Figure 8. NIST data and *k*=2 expanded uncertainty intervals at 4 MN force point for Transducer 2

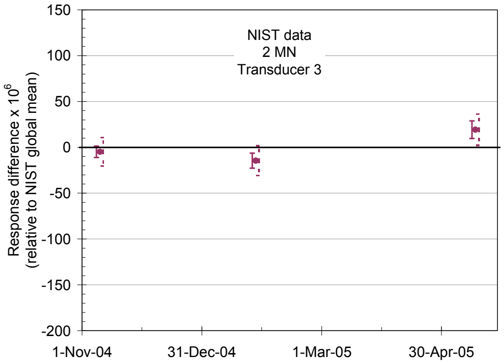

Figure 9. NIST data and *k*=2 expanded uncertainty intervals at 2 MN force point for Transducer 3

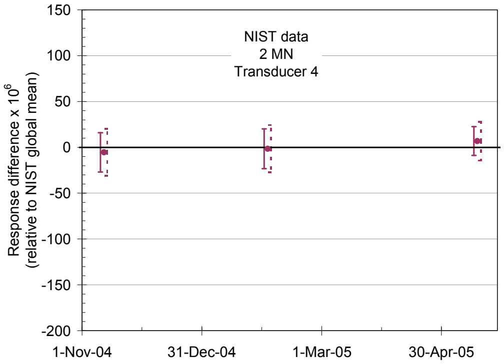

Figure 10. NIST data and *k*=2 expanded uncertainty intervals at 2 MN force point for Transducer 4

The following conclusions can be made from Fig. 5 through Fig. 10:

1. There exists no significant drift with time for any of the transducers.

2. Of the two transducers circulated for CCM.F-K4.a, Transducer 2 (Figures 7 and 8) is on the order of ten times "noisier" than Transducer 1 (Figures 5 and 6).

3. The differences among data points for Transducer 1 and Transducer 3 are larger than can be accounted for from the data-based uncertainty alone; thus there may be some source of variability associated with these transducers that is not completely addressed by the measurement statistics.

Comparison of Participating Laboratories

A graphical representation of the comparisons of all of the participating laboratories is given in Fig. 11 through Fig. 16. For each comparison, the laboratories are identified by the same lab numbers that were used in Table 2 through Table 4.

The plots for Fig. 11 through Fig. 16 are similar to those of Fig. 5 through Fig. 10, except that each ordinate now represents the difference between a particular laboratory's response and the "mean NIST pair response" (= the mean of the two NIST measurement sets immediately preceding and succeeding the laboratory's measurement set). This differencing is done in order to realize the intention of the "star" circulation for the transducers that was chosen by the CCM Force Working Group. For ease of comparison of the results for Transducers 1 and 2, Figures 11 through 14 have the same vertical scale (-600 to +600, times 10^6 relative to the NIST global mean). Figures 15 and 16, depicting the results for Transducers 3 and 4, have an expanded scale.

If k is used to indicate the lab number, then for $k \neq 1$ the difference between the response of Lab k and the corresponding mean NIST pair response is

$$d_k = r_k - (r_{k\mathrm{NIST}a} + r_{k\mathrm{NIST}b})/2 \ , \qquad (7)$$

where r_k represents the k^{th} lab mean as computed in Eq.(1) for the data set obtained by Lab k, $r_{k\mathrm{NIST}a}$ is given by Eq.(1) for the NIST data set preceding Lab k, and $r_{k\mathrm{NIST}b}$ is given by Eq.(1) for the NIST data set succeeding Lab k. For $k = 1$, designating the pilot, Lab 1, the difference is defined to be $d_1 = 0$.

The ordinates in Fig. 11 through Fig. 16, D_k, are the response differences relative to the NIST global mean, such that $D_k = (10^6)d_k/R_{\mathrm{NIST}}$, where R_{NIST} is given by Eq.(6).

In the same manner as given in Fig. 5 through Fig. 10, two expanded uncertainty intervals are shown for each point in Fig. 11 through Fig. 16. For $k \neq 1$, each left-side (solid line) bar represents the data-based expanded uncertainty $U_{ak} = 2u_{ak}$ for the corresponding data set k, where u_{ak} is calculated from Eq.(3). Each right-side (dashed line) bar represents the total expanded uncertainty $U_{tk} = 2u_{ck}$ for the corresponding data set k where u_{ck} is calculated from Eq.(4). The uncertainty bars are plotted in the figures as relative to the NIST global mean, and thus have lengths of $(10^6)U_k/R_{\mathrm{NIST}}$.

For $k = 1$, it was desired to arrive at values U_{a1} and U_{t1} for the pilot laboratory that were most comparable to U_{ak} and U_{tk} for the other laboratories. Thus U_{a1} for each of these figures is taken to be the average data-based expanded uncertainty for the NIST data sets making up the comparison: $U_{a1} = (2/m)\sum u_{aj}$, where the index j represents only the NIST measurement sets, u_{aj} is calculated from Eq.(3), and the number of NIST data sets, m, is 7 for CCM.F-K4.a and 3 for CCM.F-K4.b. The total expanded uncertainty U_{t1} is calculated similarly from the u_{cj} given by Eq.(4).

The left-side data-based uncertainty intervals indicate whether the measurement protocol is sufficient to discern differences among laboratories for a particular transducer. The total uncertainty intervals to the right of each data point indicate the significance of the differences among laboratories in light of all relevant uncertainty components – in particular, the declared uncertainties in the forces applied by the participating laboratories.

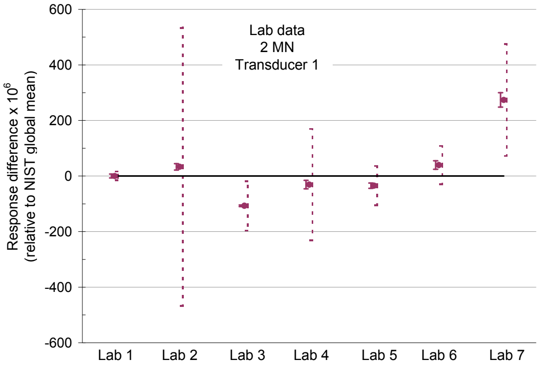

Figure 11. CCM.F-K4.a data and k=2 expanded uncertainty intervals at 2 MN force point for Transducer 1

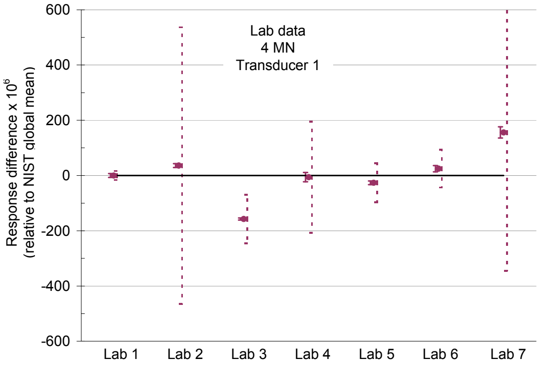

Figure 12. CCM.F-K4.a data and k=2 expanded uncertainty intervals at 4 MN force point for Transducer 1

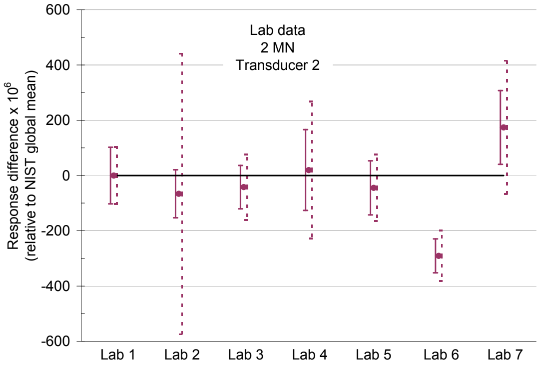

Figure 13. CCM.F-K4.a data and k=2 expanded uncertainty intervals at 2 MN force point for Transducer 2

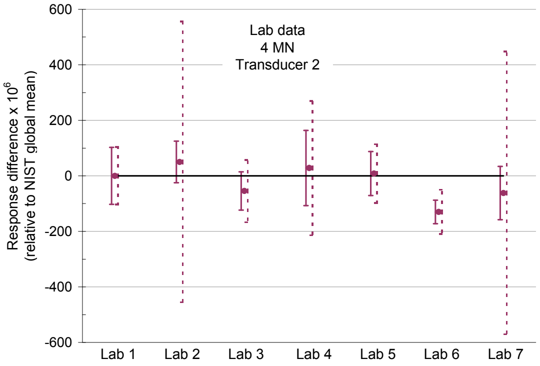

Figure 14. CCM.F-K4.a data and k=2 expanded uncertainty intervals at 4 MN force point for Transducer 2

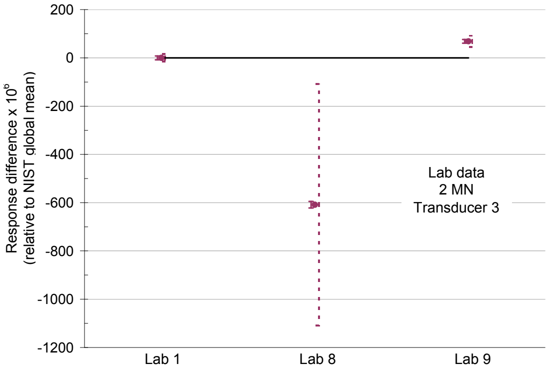

Figure 15. CCM.F-K4.b data and k=2 expanded uncertainty intervals at 2 MN force point for Transducer 3

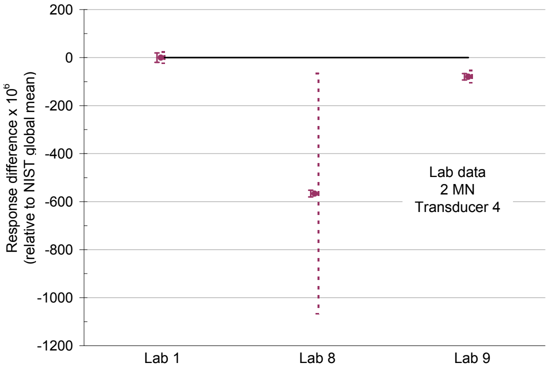

Figure 16. CCM.F-K4.b data and k=2 expanded uncertainty intervals at 2 MN force point for Transducer 4

The following conclusions can be made from Fig. 11 through Fig. 16:

1. Transducers 1, 3, and 4 appear to be capable, under the measurement protocol employed, of resolving the differences among laboratories shown in the figures. Due to its excessive variation, Transducer 2 may be of limited use in yielding significant values for these differences.

2. Based on Fig. 11, Fig. 12, Fig. 15, and Fig. 16, the results for Lab 3 of CCM.F-K4.a and Lab 8 of CCM.F-K4.b are significantly below the results of the pilot Lab 1, and possibly below estimates of the key comparison reference value, by an amount that is not accounted for by known sources of uncertainty. It is also noted that the result for Lab 7 is significantly above the pilot lab result at the 2 MN force point.

3. There are some anomalies, such as the differences in the results for Lab 6 between Transducers 1 and 2 for both 2 MN (Fig. 11 vs Fig. 13) and 4 MN (Fig. 12 vs Fig. 14). Similarly, the relative differences between Labs 1 and 9 are not consistent for Transducers 3 and 4 (Fig. 15 vs Fig. 16). These anomalies may indicate a transducer-related variability that is not yet accounted for. Normally, this variability should be accounted for by an increase in uncertainty to ensure consistency between transducers. This increase has not been applied in this report; however, this has no significant effect on the final result of the comparison.

Analysis Approach 1: "Classical" Statistical Estimation of KCRV Values and Equivalence Matrix Components

This section presents an assessment of laboratory equivalence by means of classical statistical calculations of the equivalence matrices and the Key Comparison Reference Values (KCRV) separately for each transducer.

Equivalence Matrices for Individual Transducers

The equivalence matrices are given in Table 5 through Table 8 for the 2 MN and 4 MN force points for Transducer 1 and Transducer 2, and in Table 9 for the 2 MN force point for Transducer 3 and Transducer 4. Because of the “star” nature of the circulation of the transducers, the laboratory differences given in these matrices are derived from the values of d calculated according to Eq.(7) from the response of each laboratory and its corresponding NIST pair mean response.

The “lab deltas” are denoted on the left side of each table by Δ_{kj} , where $k > j$ and k and j represent the column and row numbers, respectively, in the tables.

$$\Delta_{kj} = [d_k - d_j] , \tag{8}$$

where d_k and d_j are given by Eq.(7) for the data sets obtained by Lab k and Lab j, respectively, when $j \neq 1$. For $j = 1$, indicating the pilot lab NIST, $d_1 = 0$. The entries in the tables are multiplied by the factor $(10^6)/R_{\mathrm{NIST}}$, where R_{NIST} is given by Eq.(6), in order to present them relative to the NIST global mean.

The estimated standard deviation in the lab deltas, given on the right side of each table, is calculated for $j \neq 1$ as

$$s_{\Delta kj} = (s_k^2/n_k + s_j^2/n_j)^{1/2} \ , \qquad (9)$$

where s_k and s_j are given by Eq.(2) for the data sets obtained by Labs k and j, respectively, and both n_k and n_j equal 12. The entries in the tables are multiplied by the factor $(10^6)/R_{\mathrm{NIST}}$.

For $j = 1$, indicating the "Lab 1" row (for the pilot lab NIST) in the tables, the estimated standard deviation in the lab deltas, $s_{\Delta k1}$, is given by Eq.(9) where s_k is given by Eq.(2) for the data set obtained by Lab k, and s_j is now given by Eq.(2) for the combined data from the two NIST data sets preceding and succeeding the data set for Lab k, n_k=12, and n_j=24.

Also given in the tables are the values for Student's t-statistic for each laboratory k, calculated as $|\Delta_{kj}/s_{\Delta kj}|$. The t-statistic is shaded for all values greater than or equal to 2, indicating those laboratory pairs for which the difference between the laboratory results is statistically significant.

Table 5. Equivalence Matrix for CCM.F-K4.a, 2 MN, Transducer 1

2MN T1	Δ_{kj} x 10^6 (relative to NIST mean)						Std. dev. in Δ_{kj} x 10^6 (relative to NIST mean)						t-statistic (relative to NIST mean)					
Lab	2	3	4	5	6	7	2	3	4	5	6	7	2	3	4	5	6	7
1	33	-107	-31	-35	39	274	7	3	8	6	8	13	4.9	31.6	3.7	5.9	4.9	20.8
2		-140	-64	-68	6	241		6	10	7	10	14		23.5	6.7	9.1	0.7	16.9
3			76	72	146	381			8	5	8	13			9.9	14.7	19.0	29.1
4				-4	70	305				9	11	15				0.4	6.5	20.2
5					74	309					9	14					8.3	22.3
6						235						15						15.6
7																		

Table 6. Equivalence Matrix for CCM.F-K4.a, 4 MN, Transducer 1

4MN T1	Δ_{kj} x 10^6 (relative to NIST mean)						Std. dev. in Δ_{kj} x 10^6 (relative to NIST mean)						t-statistic (relative to NIST mean)					
Lab	2	3	4	5	6	7	2	3	4	5	6	7	2	3	4	5	6	7
1	36	-158	-6	-26	24	156	4	4	9	4	6	10	8.5	39.6	0.7	5.9	4.0	15.1
2		-194	-42	-62	-12	120		4	9	5	7	11		47.0	4.5	12.8	1.7	11.2
3			152	131	182	313			9	4	6	10			17.4	34.5	30.6	30.4
4				-20	30	162				9	10	13				2.2	3.0	12.2
5					51	182					6	11					7.8	17.1
6						131						12						11.4
7																		

Table 7. Equivalence Matrix for CCM.F-K4.a, 2 MN, Transducer 2

2MN T2	Δ_{kj} x 10^6 (relative to NIST mean)						Std. dev. in Δ_{kj} x 10^6 (relative to NIST mean)						t-statistic (relative to NIST mean)					
Lab	2	3	4	5	6	7	2	3	4	5	6	7	2	3	4	5	6	7
1	-66	-42	20	-45	-291	174	62	53	79	59	45	75	1.1	0.8	0.2	0.8	6.5	2.3
2		24	86	21	-225	240		59	85	65	53	80		0.4	1.0	0.3	4.2	3.0
3			62	-3	-249	216			83	63	50	77			0.7	0.0	5.0	2.8
4				-64	-310	154				88	79	99				0.7	3.9	1.6
5					-246	219					58	83					4.3	2.6
6						465						73						6.3
7																		

Table 8. Equivalence Matrix for CCM.F-K4.a, 4 MN, Transducer 2

4MN T2	Δ_{kj} x 10^6 (relative to NIST mean)						Std. dev. in Δ_{kj} x 10^6 (relative to NIST mean)						t-statistic (relative to NIST mean)					
Lab	2	3	4	5	6	7	2	3	4	5	6	7	2	3	4	5	6	7
1	50	-54	28	8	-130	-62	56	51	76	52	38	59	0.9	1.1	0.4	0.2	3.4	1.1
2		-105	-22	-42	-180	-112		51	77	55	43	61		2.1	0.3	0.8	4.2	1.8
3			83	63	-76	-8			76	52	40	59			1.1	1.2	1.9	0.1
4				-20	-158	-90				78	71	83				0.3	2.2	1.1
5					-138	-70					45	62					3.1	1.1
6						68						52						1.3
7																		

Table 9. Equivalence Matrices for CCM.F-K4.b, 2 MN, Transducers 3 and 4

2MN T3	Δ_{kj} x 10^6 (relative)		Std. dev. in Δ_{kj} x 10^6		t - statistic (relative)	
Lab	8	9	8	9	8	9
1	-609	69	7	6	82.8	11.5
8		677		8		86.5
9						

2MN T4	Δ_{kj} x 10^6 (relative)		Std. dev. in Δ_{kj} x 10^6		t - statistic (relative)	
Lab	8	9	8	9	8	9
1	-566	-80	10	9	55.2	8.6
8		487		10		50.6
9						

Note that for Transducer 1 about 90 % of the entries for Δ_{kj} in Table 5 and Table 6 exceed their corresponding estimated standard deviations (k=1 uncertainties) by more than a factor of 2, indicating statistical significance which implies that, under the measurement procedure employed, this transducer is capable of distinguishing differences among laboratories. The same conclusion can be made for Transducer 3 and Transducer 4 from the entries in Table 9. On the other hand, for Transducer 2, about 36 % of the entries for Δ_{kj} in Table 7 and Table 8 are greater than twice their corresponding standard deviations; thus many of these pairwise laboratory differences are not statistically significant.

A classical analysis of variance (ANOVA) was calculated from the individual responses r_i, defined in the paragraph for Eq.(1), from the data sets for each of the participants of CCM.F-K4.a. This analysis, which tests for equivalence among the data set mean values, was carried out separately for the two force points 2 MN and 4 MN for both Transducer 1 and Transducer 2. For Transducer 1, the resulting F-statistics are 233.1 and 177.1 for the 2 MN and 4 MN forces, respectively. For Transducer 2, the F-statistics are 4.29 and 1.08 for the 2 MN and 4 MN forces, respectively. The corresponding F cumulative distribution function indicates that, for a significance level of 0.05, the participants are equivalent with respect to laboratory bias only at the 4 MN force point for Transducer 2. They are not equivalent for the other three cases.

Similarly, a Levene test was carried out for the same participant data sets to test for equivalence among the data set variances. For Transducer 1, the resulting Levene F-test statistics are 5.75 and 3.53 for the 2 MN and 4 MN forces, respectively. For Transducer 2, these statistics are 1.48 and 2.22 for the 2 MN and 4 MN forces, respectively. These values yield the conclusion that the participants are equivalent with respect to laboratory variability only at the 2 MN force point for Transducer 2. They are not equivalent for the other three cases.

Key Comparison Reference Values (KCRV) based on Comparison Data Alone

Attempts are often made to determine a key comparison reference value from data obtained in key comparisons. Such a value would serve to shift the horizontal baseline used to compare laboratory results, which was arbitrarily positioned at the value of the NIST global mean, with an ordinate of 0, in Fig. 11 through Fig. 16.

A consensus mean analysis was conducted on the combined data acquired by the participating institutes, using algorithms provided by the DATAPLOT software system for scientific statistical analysis, available from the NIST Statistical Engineering Division [4]. Documentation and procedures for acquiring this software are available at the Internet address http://www.itl.nist.gov/div898/software/dataplot/homepage.htm. The consensus mean analysis computes estimates of the consensus mean, and the associated uncertainties, based on all of the comparison data using a variety of methods [5-9].

Separate estimates of uncertainty of the applied forces reported by the participants do not enter into this analysis.

The values of the consensus mean and associated expanded uncertainty (k=2) are provided in Table 10 for several analysis methods for the 2 MN and 4 MN force points employed for Transducer 1 and Transducer 2. The results of the computations are given in Table 10 in the unit (mV/V) of the indicating instrument used to acquire the transducer responses.

Table 10. Consensus Mean Analysis (indicator unit of mV/V)

2 MN, Transducer 1		4 MN, Transducer 1		2 MN, Transducer 2		4 MN, Transducer 2		
Consensus Mean (mV/V)	Expanded Uncertainty k = 2 (mV/V)	Consensus Mean (mV/V)	Expanded Uncertainty k = 2 (mV/V)	Consensus Mean (mV/V)	Expanded Uncertainty k = 2 (mV/V)	Consensus Mean (mV/V)	Expanded Uncertainty k = 2 (mV/V)	Analysis Method
0.799 209	0.000 068	1.598 721	0.000 106	0.999 500	0.000 101	1.999 924	0.000 090	Mandel-Paule
0.799 209	0.000 068	1.598 721	0.000 106	0.999 499	0.000 101	1.999 922	0.000 091	Modified M.Paule
0.799 209	0.000 068	1.598 721	0.000 106	0.999 499	0.000 101	1.999 925	0.000 089	Vangel-Rukhin ML
0.799 209	0.000 181	1.598 721	0.000 299	0.999 507	0.000 282	1.999 941	0.000 203	Bound on Bias
0.799 206	0.000 213	1.598 717	0.000 275	0.999 470	0.000 379	1.999 881	0.000 366	Schiller-Eberhardt
0.799 209	0.000 074	1.598 721	0.000 115	0.999 507	0.000 110	1.999 941	0.000 085	Mean of Means
0.799 149	0.000 002	1.598 651	0.000 004	0.999 480	0.000 029	1.999 901	0.000 050	Graybill-Deal
0.799 200	0.000 016	1.598 719	0.000 024	0.999 522	0.000 023	1.999 959	0.000 018	Grand Mean
0.799 209	0.000 091	1.598 720	0.000 142	0.999 500	0.000 135	1.999 925	0.000 109	Generalized CI
0.799 208	0.000 073	1.598 721	0.000 115	0.999 501	0.000 109	1.999 928	0.000 094	DerSimonian-Laird
0.799 190		1.598 716		0.999 540		1.999 980		NIST mean

Regarding the recommendation of a single KCRV value, parsimony suggests the choice of the mean of means statistic -- this is highlighted in Table 10 via the shaded row. Thus, from Table 10, a summary of the four recommended values is as follows:

Force (MN)	Transducer	KCRV (mV/V)	K=2 Exp.Unc. (mV/V)	K=2 Rel.Exp. Unc. (%)
2	1	0.799209	0.000074	0.0093
4	1	1.598721	0.000115	0.0072
2	2	0.999507	0.000110	0.0110
4	2	1.999941	0.000085	0.0042

Table 11 provides the same results as a relative difference with respect to the mean of means value; specifically, each relative consensus mean value in Table 11 is computed from the corresponding value in Table 10 by: 10^6 x [(consensus mean) – (mean of means)] / (mean of means), where (mean of means) denotes the consensus mean value by the mean of means method. Each relative expanded uncertainty value in Table 11 is computed from the corresponding value in Table 10 by: 10^6 x (expanded uncertainty) / (mean of means), where, again, (mean of means) denotes the consensus mean value by the mean of means method.

Table 11. Consensus Mean Analysis (relative to Mean of Means x 10^6)

2 MN, Transducer 1		4 MN, Transducer 1		2 MN, Transducer 2		4 MN, Transducer 2		
Consensus Mean - Mean of Means (relative)	Expanded Uncertainty k = 2 (relative)	Consensus Mean - Mean of Means (relative)	Expanded Uncertainty k = 2 (relative)	Consensus Mean - Mean of Means (relative)	Expanded Uncertainty k = 2 (relative)	Consensus Mean - Mean of Means (relative)	Expanded Uncertainty k = 2 (relative)	Analysis Method
0	85	0	66	-6	101	-8	45	Mandel-Paule
-1	85	0	66	-7	101	-9	46	Modified M.Paule
-1	85	0	66	-8	101	-8	44	Vangel-Rukhin ML
0	227	0	187	0	282	0	102	Bound on Bias
-4	267	-3	172	-36	379	-30	183	Schiller-Eberhardt
0	93	0	72	0	110	0	42	Mean of Means
-76	2	-44	2	-27	29	-20	25	Graybill-Deal
-11	20	-1	15	15	23	9	9	Grand Mean
-1	114	0	89	-7	135	-8	55	Generalized CI
-1	92	0	72	-6	109	-6	47	DerSimonian-Laird
-24		-3		33		20		NIST mean

This analysis was not conducted for Transducer 3 and Transducer 4 because of the very small population sample for comparison CCM.F-K4.b.

The "mean of means" value yielded by the consensus mean analysis is the same as the lab mean calculated from

$$R_{\mathrm{LABS}} = (1/m)\sum r_k \,, \qquad (10)$$

where the index k indicates the lab number and the number of participating laboratories, m, is 7 for CCM.F-K4.a and 3 for CCM.F.K4.b. For $k \neq 1$, r_k is the mean of the twelve observations from the data set acquired by Lab k, as given by Eq.(1). For k=1, r_1 is the NIST global mean given by Eq.(6) from a total of 7x12=84 observations for CCM.F-K4.a and 3x12=36 observations for CCM.F-K4.b.

It is seen from Table 11 that six of the ten consensus mean values lie within 0.0010 % of the mean of means values, for each transducer and force point. Only one method, of Graybill-Deal, yields values differing from the mean of means values by more than 0.0015 % for Transducer 1, and only the Schiller-Eberhardt and Graybill-Deal methods yield values differing from the mean of means values by more than 0.0015 % for Transducer 2.

KCRV Calculated from Lab – NIST Differences and Uncertainty Weightings

Possible quantities often proposed as candidates for a key comparison reference value include the unweighted mean, the weighted mean, and the median of the participating laboratory results. For the comparisons that have been presented here, the participant results consist of the differences d_k among the laboratories, given by Eq.(7) with the stipulation that $d_1=0$. The unweighted mean is then calculated as

$$V = (1/m)\sum d_k \,, \tag{11}$$

and the median value is the median of the set $[d_1, \ldots, d_m]$, where m is the number of participants.

If each value of d_k has a corresponding standard uncertainty u_k, a weighted mean may be calculated as

$$W = \sum(d_k/u_k^2)/\sum(1/u_k^2). \tag{12}$$

Two values of the weighted mean are calculated, corresponding to the "data-based" uncertainties and the "total" uncertainties represented by the two sets of uncertainty intervals depicted in Fig. 11 through Fig. 16. For the weighted mean corresponding to the "data-based" uncertainties, the u_k in Eq.(12) is calculated from Eq.(3), using comparison measurement data only. For the weighted mean corresponding to the "total" uncertainties, the u_k in Eq.(12) is the combined standard uncertainty calculated from Eq.(4), incorporating the additional uncertainties associated with the applied forces and the measuring amplifier corrections as discussed in the paragraph preceding Eq.(4).

Table 12 and Table 13 give the four computations of the key comparison reference values, for the two force points and the four transducers used, with the values given in the unit (mV/V) of the indicating instruments in Table 12 and in relative units in Table 13 by multiplying by $10^6/R_{\mathrm{NIST}}$.

An additional entry is given at the bottom of these tables, which gives values for the difference between the mean of means from Eq.(10) and the NIST global mean from Eq.(6). These values would correspond to the unweighted means if the d_k were not calculated from Eq.(7), but simply from $d_k=r_k-R_{\mathrm{NIST}}$ (thus ignoring the star circulation of the comparison). The tables show that the (mean of means – NIST global mean) values differ from the unweighted mean values by no more than 0.0003 %. This implies that transducer drift is not significantly apparent in the comparison data.

Table 12. Key Comparison Reference Values (indicator unit of mV/V)

2 MN, Transducer 1	4 MN, Transducer 1	2 MN, Transducer 2	4 MN, Transducer 2	2 MN, Transducer 3	2 MN, Transducer 4	Description
0.000 020	0.000 006	-0.000 036	-0.000 046	-0.000 357	-0.000 388	unweighted mean
-0.000 001	-0.000 007	-0.000 097	-0.000 107	0.000 043	-0.000 068	weighted mean (using "total" uncertainty)
-0.000 059	-0.000 105	-0.000 101	-0.000 116	-0.000 105	-0.000 445	weighted mean (using "data-based" uncertainty)
0.000 000	0.000 000	-0.000 042	0.000 000	0.000 000	-0.000 144	median
0.000 019	0.000 005	-0.000 033	-0.000 039	-0.000 362	-0.000 389	(Mean of means - NIST global mean)

Table 13. Key Comparison Reference Values (relative to NIST global mean x 10^6)

2 MN, Transducer 1	4 MN, Transducer 1	2 MN, Transducer 2	4 MN, Transducer 2	2 MN, Transducer 3	2 MN, Transducer 4	Description
25	4	-36	-23	-180	-215	unweighted mean
-1	-4	-97	-54	22	-38	weighted mean (using "total" uncertainty)
-74	-66	-101	-58	-53	-247	weighted mean (using "data-based" uncertainty)
0	0	-42	0	0	-80	median
24	3	-33	-20	-183	-216	(Mean of means - NIST global mean)

The key comparison reference values given in these tables show a range, over the four methods used, of about 0.01 % for Transducer 1 to over 0.02 % for Transducer 3 and Transducer 4. The range is attributable largely to the influence of the uncertainty on the weighted mean values. Because of the large variation in the uncertainty of the applied forces reported by the participants, it may not be possible to select meaningful key comparison reference values.

If it is desired to have such reference values as a product of these comparisons, the unweighted mean, given in the first line of Table 12 and Table 13, may be the most reasonable choice, because (a) it is less affected by “outside factors”, (b) it may be less affected by large variations in the results when the population is small, and (c) it corresponds most closely to the values yielded by the consensus means analysis. Because the unweighted mean is essentially the same as the mean of means value from the consensus means analysis, the expanded uncertainty yielded by that analysis for the mean of means could be used as a reasonable estimate for the expanded uncertainty in the unweighted mean key comparison reference value.

Graphical Representation of Participant Results with Candidate KCRVs

A graphical representation of the comparisons of all of the participating institutes is given in Fig. 17 through Fig. 22. The baseline in these figures has been arbitrarily chosen to be the unweighted mean KCRV, V, given by Eq.(11). Thus the ordinates in these figures, D_k, are the response differences d_k, offset from this baseline, and scaled to be relative to the NIST mean R_{NIST} given Eq.(6): $D_k = (10^6)(d_k\text{-}V)/R_{\mathrm{NIST}}$. Because of the 10^6 multiplier, the ordinates can informally be regarded to be in "parts/million".

The unweighted mean baselines in Fig.17 through Fig.20 can also be regarded to represent the mean of means values given by the shaded lines in Tables 10 and 11. Thus the expanded uncertainties, given in values relative to the NIST mean x 10^6 in Table 11, are shown to represent the expanded uncertainty intervals about the baselines, appearing as bands indicated the heavy dashed horizontal lines, in Fig.17 through Fig.20. These expanded uncertainty intervals are not given for Fig.21 and Fig.22, because, as indicated below Table 11, the Consensus Mean Analysis was not conducted for Transducers 3 and 4.

Two expanded uncertainty intervals are shown for each point in Fig. 17 through Fig. 22, representing "data-based" expanded uncertainties (to the left of each point) calculated from the comparison data alone, and "total" expanded uncertainties (to the right of each point) that incorporate other significant uncertainty contributors. With the exception of the pilot lab NIST, each left-side (solid line) bar represents the data-based expanded uncertainty $U_{ak} = 2u_{ak}$ for the corresponding data set k, where $k \neq 1$ and u_{ak} is calculated from Eq.(3). Each corresponding right-side (dashed line) bar represents the total expanded uncertainty $U_{tk} = 2u_{ck}$ for the corresponding data set k where u_{ck} is calculated from Eq.(4). The uncertainty bars are plotted in the figures as relative to the NIST mean, and thus have lengths of $(10^6)U_k/R_{\mathrm{NIST}}$.

For the pilot lab NIST ($k = 1$), it was desired to arrive at values U_{a1} and U_{t1} that were most comparable to U_{ak} and U_{tk} for the other laboratories. Thus U_{a1} for each of these figures is taken to be the average data-based expanded uncertainty for the NIST data sets making up the comparison: $U_{a1} = (2/m)\sum u_{aj}$, where the index j represents only the NIST measurement sets, u_{aj} is calculated from Eq.(3), and the number of NIST data sets, m, is 7 for CCM.F-K4.a and 3 for CCM.F-K4.b. The total expanded uncertainty U_{t1} is calculated similarly from the u_{cj} given by Eq.(4).

The left-side data-based uncertainty intervals indicate whether the measurement protocol is sufficient to discern differences among laboratories for a particular transducer. The total uncertainty intervals to the right of each data point indicate the significance of the differences among laboratories in light of all relevant uncertainty components – in particular, the declared uncertainties in the forces applied by the participating laboratories.

The vertical scale in the figures has been chosen maximize the detail of the data-based uncertainty intervals while keeping the data points within the bounds of the plots, at the expense of allowing some of the total uncertainty intervals to extend out of the range.

The values of the four key comparison reference values given in Table 13, designated as the unweighted mean, the weighted mean using “total” uncertainty (labeled as “total unc” in the plot legends), the weighted mean using “data-based” uncertainty (labeled as “data unc” in the legends), and the median, are shown as horizontal lines in Fig. 17 through Fig. 22. Each of these values has been offset from the baseline, and scaled by the factor $10^6/R_{\mathrm{LABS}}$, to be comparable with the points on the plots. Thus the unweighted mean lies at ordinate 0.

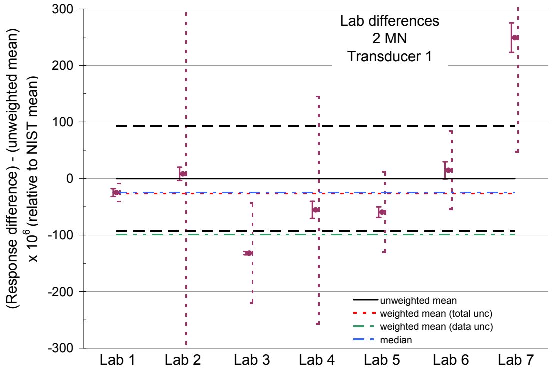

Figure 17. CCM.F-K4.a data, *k*=2 expanded uncertainties, and KCRVs at 2 MN force point for Transducer 1

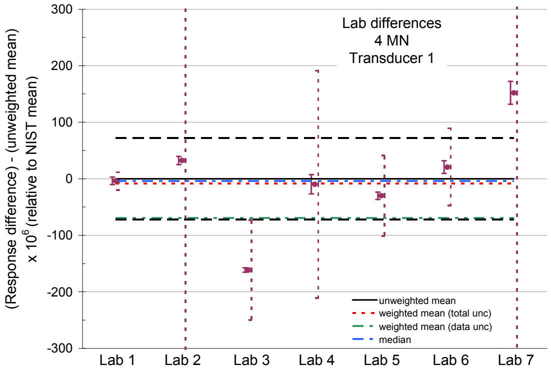

Figure 18. CCM.F-K4.a data, *k*=2 expanded uncertainties, and KCRVs at 4 MN force point for Transducer 1

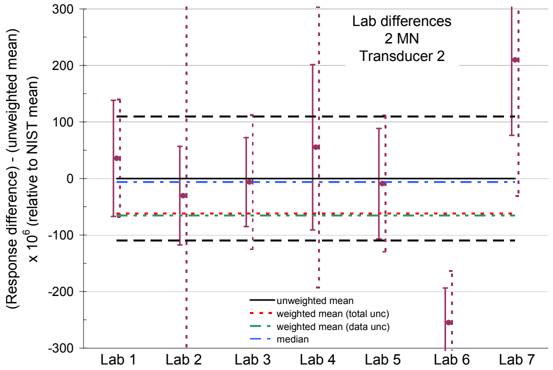

Figure 19. CCM.F-K4.a data, *k*=2 expanded uncertainties, and KCRVs at 2 MN force point for Transducer 2

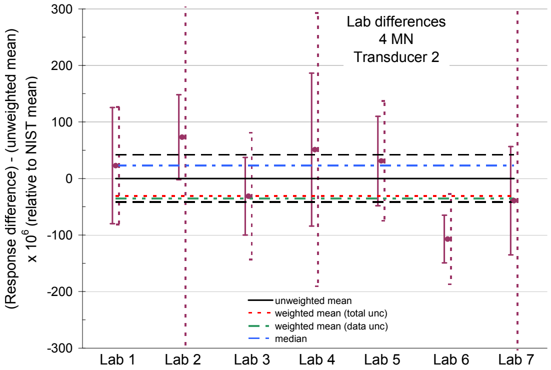

Figure 20. CCM.F-K4.a data, *k*=2 expanded uncertainties, and KCRVs at 4 MN force point for Transducer 2

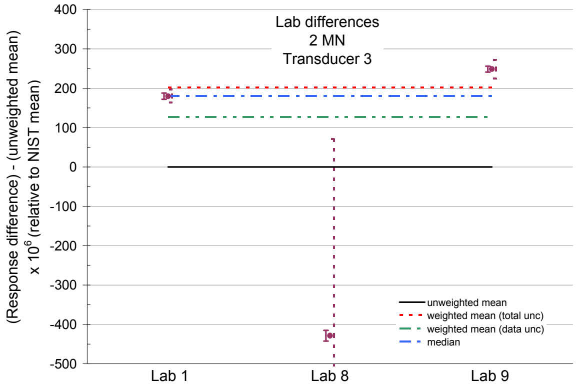

Figure 21. CCM.F-K4.b data, *k*=2 expanded uncertainties, and KCRVs at 2 MN force point for Transducer 3

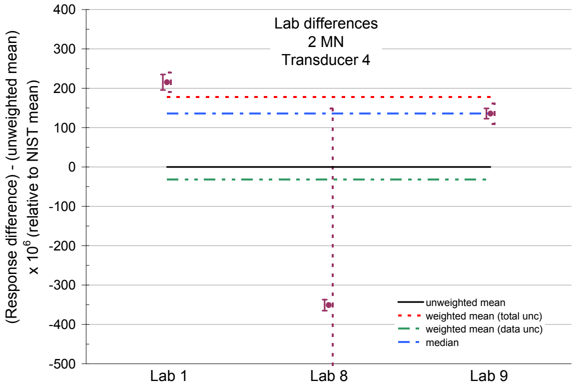

Figure 22. CCM.F-K4.b data, *k*=2 expanded uncertainties, and KCRVs at 2 MN force point for Transducer 4

Conclusions from Approach 1

1. Transducers 1, 3, and 4 appear to be capable, under the measurement protocol employed, of resolving the differences among laboratories shown in Fig. 17 through Fig. 22. Transducer 2, which appears to be on the order of ten times "noisier" than Transducer 1, may be of limited use in yielding significant values for these differences.

2. No significant drift with time is seen in the transducer characteristics.

3. On the basis of Transducer 1 for CCM.F-K4.a, Lab 3 is significantly below the pilot Lab 1, and possibly below other estimates of a key comparison reference value, by an amount that may not be accounted for by known sources of uncertainty. Lab 7 may be correspondingly high, especially at the 2 MN force point.

4. On the basis of Transducers 3 and 4 for CCM.F-K4.b, Lab 8 may be significantly low.

5. There are some anomalies, such as the differences in the results of Lab 6 for Transducers 1 and 2 as shown in Fig. 17 through Fig. 20, and in the relative differences between Lab 1 and Lab 9 for both Transducers 3 and 4 as seen in Fig. 21 and Fig. 22, that may indicate a transducer-related variability that is not accounted for.

6. If a key comparison reference value is desired, the unweighted mean, which is essentially given by the (mean of means) - (NIST global mean), may be the most appropriate, with the uncertainty yielded by the consensus mean analysis for the mean of means.

7. The use of the Mean of Means value as an estimator of the KCRV yields the following estimates:

Force (MN)	Transducer	KCRV (mV/V)	K=2 Exp.Unc. (mV/V)	K=2 Rel.Exp. Unc. (%)
2	1	0.799209	0.000074	0.0093
4	1	1.598721	0.000115	0.0072
2	2	0.999507	0.000110	0.0110
4	2	1.999941	0.000085	0.0042

8. The equivalence matrices shown in Table 5 through Table 8 indicate that, of the 21 paired comparisons of the 7 laboratories participating in CCM.F-K4.a, about 90 % had t-statistic values that indicated statistically significant pair differences for Transducer 1, and 36 % had t-statistics that indicated statistically significant differences for Transducer 2.

9. The equivalence matrices shown in Table 9 indicate that, of the 3 paired comparisons of the 3 laboratories participating in CCM.F-K4.b, all had t-statistic

values that indicated statistically significant pair differences for both Transducers 3 and 4.

10. Inspection of Fig.17 and Fig.18 shows that, for Transducer 1, the data-based expanded uncertainty intervals for Lab 3 and Lab 7 lie outside of the expanded uncertainty band for the baseline, which represents the mean of means value yielded by the Consensus Means Analysis. This is true for both the 2 MN and 4 MN force points.

11. Inspection of Fig.19 and Fig.20 shows that, for Transducer 2, the data-based expanded uncertainty interval for Lab 6 lies outside of the expanded uncertainty band for the baseline mean of means value. This is true for both the 2 MN and 4 MN force points.

Analysis Approach 2: Metrologia-Based Estimation of KCRV Values and Equivalence Matrix Components

The CCM Force working Group, at the meeting at CENAM in December, 2007, expressed a desire to apply a consistent analysis of the comparison data for each of the four pair of key comparisons in force. Specifically, an objective was to obtain a comparison at each force point based on the combined data from the transducers employed for that force point, using a weighted mean calculation of the KCRV.

This appendix presents such an analysis for the 2 MN and 4 MN force points based on the combined data from Transducers 1 through 4 at 2 MN, and from Transducers 1 and 2 at 4 MN.

Uncertainty Associated with Transducer Variability Estimated from NIST Data

The comparison of the data for each transducer from the repeated measurement sets at NIST, as presented in Fig. 5 through Fig. 10, indicate a variation in the responses of Transducers 1 and 3 that is not accounted for by the uncertainties in the measurement sets given by Eq.(3). This variation has an apparent random character as opposed to a significant long-term slope normally referred to as drift. While the cause of this variation is not known, a standard uncertainty term, u_x, can be ascribed to this effect and chosen to be large enough for a consistency check of the NIST results for a particular transducer to pass. This is done in the following manner.

If the differences, d_j, in the individual NIST measurement set results from the NIST mean for a particular transducer are given by Eq.(5), the combined standard uncertainty in each d_j is calculated from

$$u_{dj}^2 = u_{aj}^2 + u_v^2 + u_x^2 \ , \qquad (13)$$

where u_{aj} is the standard uncertainty calculated from Eq.(3) for the data points of the NIST measurement set j, u_v is the standard uncertainty in the measuring amplifier corrections, and u_x is the standard uncertainty associated with the source of variability that is not accounted for by the measurement set statistics embodied by the u_{aj}. u_x can be considered to relate to an apparent "long term" variation in the transducer response, where "long term" relates to the period of time between successive measurement sets, typically a few months.

Note that u_f is not included here because all of these measurement sets were acquired with the same deadweight force standard, and any error in the applied force applies equally to all j measurement sets.

u_x is determined by first calculating a weighted average of the j measurement set results, as well as the corresponding standard uncertainty, in a manner that is consistent with the calculation of values of the weighted mean KCRV:

$$K = \sum(d_j/u_{dj}^2) \,/\, \sum(1/u_{dj}^2) \ , \tag{14}$$

$$u_K^2 = 1 \,/\, \sum(1/u_{dj}^2) \ . \tag{15}$$

The symbol K, rather than $KCRV$, is used in Eq.(14) to avoid confusion with the calculation of the weighted mean KCRVs from the participating laboratory measurement sets in Eq.(23) of the following section. The weighted mean calculation and the following consistency check are described in reference [10].

A chi-squared consistency check for the NIST measurement sets conducted at different times is performed by calculating a chi-squared value as

$$\chi^2 = \sum \left[(d_j - K)^2 \,/\, u_{dj}^2 \right] , \tag{16}$$

and then determining the chi-squared probability value, P, from

$$P = \text{CHIDIST}(\chi, j\text{-}1) \ , \tag{17}$$

where j-1 is the number of degrees of freedom.

The consistency check is regarded to fail if $P < 0.05$ [10].

The value of u_x for each transducer was determined by repeating the consistency check described above for increasing values of u_x until the consistency check was found to pass.

This procedure yielded relative values of u_x of 0.0006 % and 0.0008 % for Transducers 1 and 3, respectively, relative to the value of R_{NIST} for each transducer. The values of u_x

may remain at zero for Transducers 2 and 4. These values of u_x enable the consistency checks to pass for the NIST results for each force point with each transducer.

Weighted Mean KCRV Calculation for Individual Transducers

Before calculating reference values from the combined data of multiple transducers, the KCRV must be obtained for each transducer from the appropriate measurement sets of the participating laboratories, taking into account the additional uncertainty component u_x discussed in the previous section. In addition, a proper accounting must be taken of the uncertainties involved in the NIST force standard's function as a "Pilot Link machine."

The difference, d_k, between the response of Lab k and the corresponding NIST pair response is given by Eq.(7), with d_1 for Lab 1 (NIST) defined to be zero.

The combined standard uncertainty, u_{ck}, for the measurement set of Lab k, for $k > 1$, is calculated according to Eq.(4), taking into account the standard deviation of the measurement set data, the uncertainty in the applied force for Lab k, and the uncertainty in the measuring amplifier corrections.

In order to determine the uncertainty, u_{dk}, in the value of d_k, the uncertainty for the mean response of the two NIST measurement sets immediately preceding and succeeding the Lab k set must first be computed.

The pilot laboratory, NIST, is serving as a link among the other participating laboratories, and is also a participant itself. Because all of NIST's measurement sets were acquired with the same deadweight force standard, the uncertainty in the applied force at NIST does not enter into the calculation for u_{dk}.

Let $u_{k\mathrm{PLMa}}$ be the combined standard uncertainty of the "Pilot Link Machine" (PLM) measurement set immediately preceding the Lab k measurement set, and $u_{k\mathrm{PLMb}}$ be the combined standard uncertainty of the PLM measurement set immediately succeeding the Lab k measurement set. When the PLM measurement sets are functioning as links to the other participants, their corresponding standard uncertainties are given as

$$u_{k\mathrm{PLMa}}^2 = u_{ak\mathrm{PLMa}}^2 + u_v^2 + u_x^2 \quad , \qquad (18)$$

$$u_{k\mathrm{PLMb}}^2 = u_{ak\mathrm{PLMb}}^2 + u_v^2 + u_x^2 \quad . \qquad (19)$$

In Eqs.(18-19), $u_{ak\mathrm{PLMa}}$ and $u_{ak\mathrm{PLMb}}$ are the standard uncertainties in the measurement data in the corresponding pilot lab sets, calculated from Eq.(3). u_v is the standard uncertainty in determining the measuring amplifier corrections, and u_x is the standard uncertainty associated with transducer variability, determined through the use of the chi-squared consistency check as discussed in the previous section.

The standard uncertainty of the PLM measurement set pair used as the link to Lab k is given by

$$u_{k\mathrm{PLM}}^2 = (u_{k\mathrm{PLMa}}^2 + u_{k\mathrm{PLMb}}^2) / 2 \quad . \tag{20}$$

Finally, the standard uncertainty in the value of d_k, for $k > 1$, is given by

$$u_{dk}^2 = u_{k\mathrm{PLM}}^2 + u_{ck}^2 \quad , \tag{21}$$

where u_{ck} is given by Eq.(4).

For $k = 1$, d_1 is defined to be zero. However, d_1 does have a finite uncertainty, which corresponds to the PLM's function as a participant in the comparison. Accordingly,

$$u_{d1}^2 = [(1/m)\sum u_{j\mathrm{PLM}}^2]+ u_{f1}^2 \quad , \tag{22}$$

where the summation is over the index j from 1 to m, m=7 for Transducers 1 and 2, m=3 for Transducers 3 and 4, $u_{j\mathrm{PLM}}^2$ is computed in the manner of Eq.(20), and u_{f1} is the standard uncertainty in the applied force for the PLM.

For each force point for each transducer, a set of d_k and u_{dk}^2 are now known. A KCRV, in the form of a weighted mean of the d_k, and its corresponding standard uncertainty, can be calculated for each force point for each transducer from

$$KCRV = \sum(d_k/u_{dk}^2) / \sum(1/u_{dk}^2) \quad , \tag{23}$$

$$u_{KCRV}^2 = 1 / \sum(1/u_{dk}^2) \quad . \tag{24}$$

A chi-squared consistency test is performed in the same manner as seen in Eq.(16-17):

$$\chi^2 = \sum [(d_k - KCRV)^2 / u_{dk}^2] \quad , \tag{25}$$

$$P = \mathrm{CHIDIST}(\chi, k\text{-}1) \quad . \tag{26}$$

Figures 23 through 28 present the differences of the participant laboratory responses and the weighted mean KCRV. The ordinates are scaled to values that are relative to the NIST mean. Thus the ordinate, D_k, for a particular laboratory, Lab k, in each plot is calculated as

$$D_k = (10^6)(d_k - KCRV) / R_{\mathrm{NIST}} \quad , \tag{27}$$

where d_k is given by Eq.(7), $KCRV$ is given by Eq.(23), and R_{NIST} is given by Eq.(6). The error bars represent the expanded uncertainty intervals, also scaled to relative values, for each lab:

$$U_{Dk} = 2u_{Dk} \quad , \tag{28}$$

where

$$u_{Dk} = (10^6)u_{dk} \,/\, R_{\mathrm{NIST}} \quad , \tag{29}$$

and u_{dk} is given by Eq.(21-22). The weighted mean KCRV is the baseline at ordinate zero, and its expanded uncertainty of $2(u_{KCRV})$, where u_{KCRV} is given by Eq.(24), is depicted by the dashed horizontal lines. As all ordinates and expanded uncertainty intervals are multiplied by 10^6 in the figures, they can informally be regarded to represent "parts/million."

Of the six individual transducer weighted mean KCRV calculations represented by Fig. 23 through Fig. 28, only the comparison for 4 MN with Transducer 2 passes its consistency check. Better consistency is obtained, however, when the combined data from multiple transducers is used to calculate the KCRV, as presented in the following section.

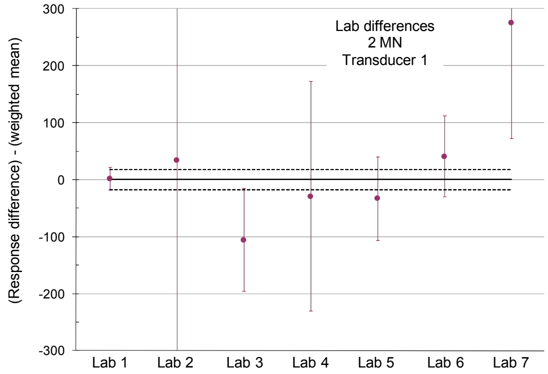

Figure 23. CCM.F-K4.a data, weighted mean KCRV, and *k*=2 uncertainties, at 2 MN force point for Transducer 1

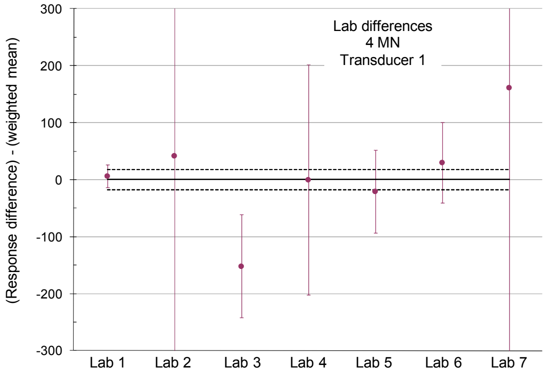

Figure 24. CCM.F-K4.a data, weighted mean KCRV, and *k*=2 uncertainties, at 4 MN force point for Transducer 1

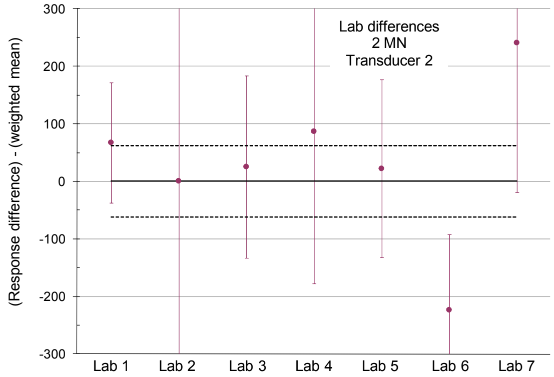

Figure 25. CCM.F-K4.a data, weighted mean KCRV, and *k*=2 uncertainties, at 2 MN force point for Transducer 2

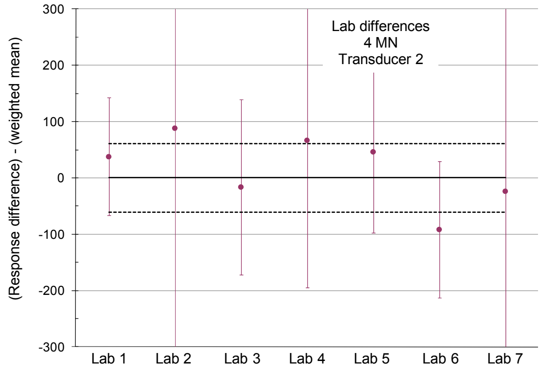

Figure 26. CCM.F-K4.a data, weighted mean KCRV, and *k*=2 uncertainties, at 4 MN force point for Transducer 2

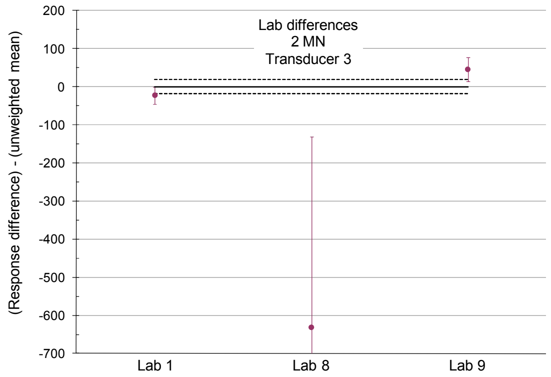

Figure 27. CCM.F-K4.b data, weighted mean KCRV, and *k*=2 uncertainties, at 2 MN force point for Transducer 3

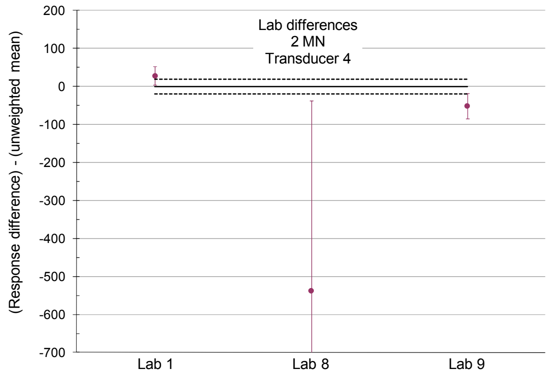

Figure 28. CCM.F-K4.b data, weighted mean KCRV, and *k*=2 uncertainties, at 2 MN force point for Transducer 4

KCRV Calculation from Combined Data for Multiple Transducers

Each participating NMI conducted measurements on two transducers that were circulated together. The comparison protocol specified that the two transducers be from different manufacturers, to reduce the possibility of the comparison results being influenced by the characteristics of a particular transducer design. It was intended that a single result for each laboratory be extracted from the combined data of both transducers in a circulating pair, for each force point in the comparison (2 MN and 4 MN for CCM.F-K4.a, and 2 MN for CCM.F-K4.b).

The objective of the key comparisons in force is to compare the forces applied by the participants' force standards. While the outputs of the transducers in each pair are of different magnitudes, each transducer output varies essentially linearly with the applied force. Thus the normalization of the results for each transducer to values that are relative to the appropriate NIST mean responses for that transducer makes an analysis of the combined results possible. Since the results derive from the differences in the transducer responses obtained at the participating institutes, the combined normalized results represent estimates of the differences in the forces applied by the participants' force standard machines.

The previous section utilized results for the participants in terms of d_k, from Eq.(7), and u_{dk}, from Eqs.(21-22). For the k^{th} participant of CCM.F-K4.a, these results were obtained for Transducers 1 and 2. Let these values be normalized to the appropriate mean NIST pair responses, where the term "mean NIST pair response" is discussed in the paragraph presenting Eq.(7). For Transducer 1, the normalized differences and standard uncertainties are

$$D_{k1} = d_{k1} / R_{k1\mathrm{NIST}} \quad \text{and} \tag{30}$$

$$u_{Dk1} = u_{dk1} / R_{k1\mathrm{NIST}} \quad , \tag{31}$$

where $R_{k1\mathrm{NIST}}$ is the mean NIST pair response for lab k for Transducer 1 given by

$$R_{k1\mathrm{NIST}} = (r_{k1\mathrm{NIST}a} + r_{k1\mathrm{NIST}b}) / 2 \quad . \tag{32}$$

D_{k2}, and u_{Dk2} are computed similarly for Transducer 2.

It is noted, however, that the figures and the equivalence matrices presented in this section are unchanged, regardless of whether the denominator in Eqs.(30-31) is the mean NIST pair response $R_{k1\mathrm{NIST}}$ given by Eq.(32), the NIST global mean R_{NIST} given by Eq.(6) for the same transducer, or the mean over all laboratories R_{LABS} given by Eq.(10) for the same transducer.

A weighted mean can be calculated from these two results, after first taking into account that both results were obtained with the same force standard machine of Lab k, such that the uncertainty in the applied force is not incorporated into the weighted mean calculation:

$$u'^2_{Dk1} = u^2_{Dk1} - u^2_{fk} \quad , \tag{33}$$

where u_{fk} is the standard uncertainty in the applied force for Lab k when expressed as a fraction of the applied force. A similar calculation is carried out for u'^2_{Dk2}.

Then the weighted mean of D_{k1} and D_{k2} is given by

$$K_k = (D_{k1}/u'^2_{Dk1} + D_{k2}/\, u'^2_{Dk2}\,) \,/\, (1/u'^2_{Dk1} + 1/\, u'^2_{Dk2}\,) \quad , \tag{34}$$

and its standard uncertainty is given by

$$u_{Kk}^2 = 1/\,(1/u'^2_{Dk1} + 1/\, u'^2_{Dk2}\,) \quad . \tag{35}$$

The uncertainty in the applied force for Lab k, u_{fk}, must now be reincorporated into the combined standard uncertainty as

$$u_{cKk}^2 = u_{Kk}^2 + u_{fk}^2 \quad . \tag{36}$$

At this point, a weighted mean, K_k, and its associated combined standard uncertainty, u_{cKk}, have been calculated for the combined data from Transducers 1 and 2, for both force points 2 MN and 4 MN for each Lab k. For each force point, a new weighted mean can be calculated from K_k and u_{cKk} for each force point over the 7 laboratories of CCM.F-K4.a, using the form of Eqs.(23-24).

In the same manner, a new weighted mean can be calculated from K_k and u_{cKk} for Transducers 3 and 4 for each force point over the 3 laboratories of CCM.F-K4.b, again using the form of Eqs.(23-24).

The differences of the participant laboratory responses from the new weighted means calculated from the combined results of Transducers 1 and 2 are shown in Fig. 29 for 2 MN and in Fig. 30 for 4 MN. The differences of the participant laboratory responses from the new weighted means calculated from the combined results of Transducers 3 and 4 are shown in Fig. 31 for 2 MN. The scaling and expanded uncertainty interval representations are the same as for Fig. 23 through Fig. 28.

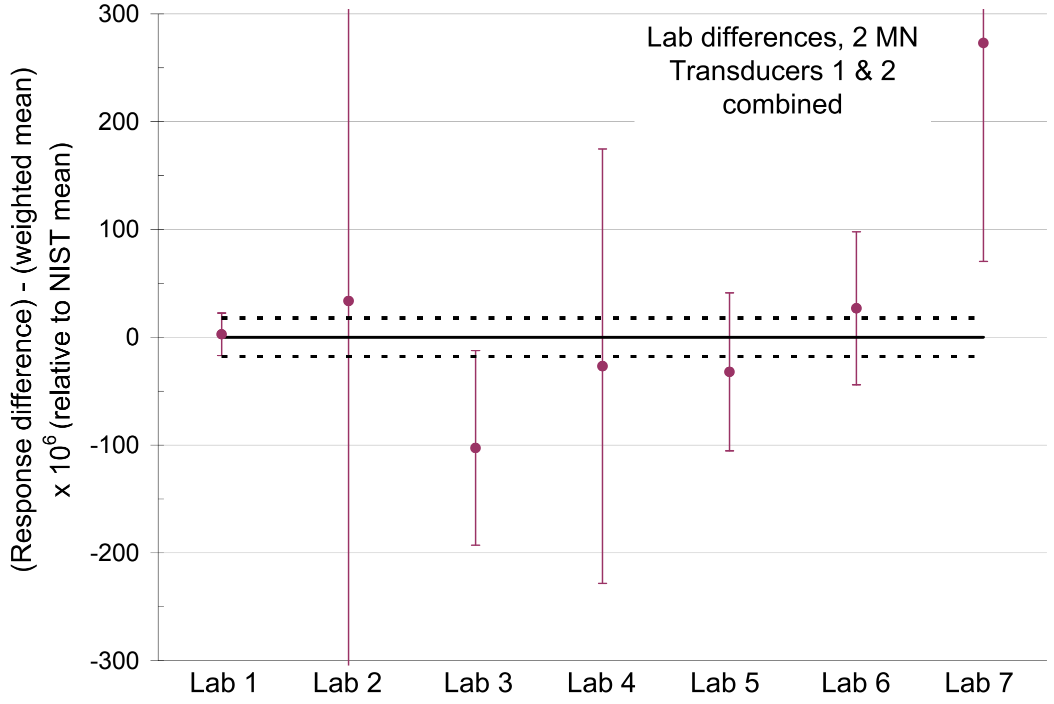

Figure 29. Differences between participant laboratory responses and the weighted mean KCRV, with k=2 expanded uncertainty intervals, calculated from the combined data from Transducers 1 and 2 at the 2 MN force point

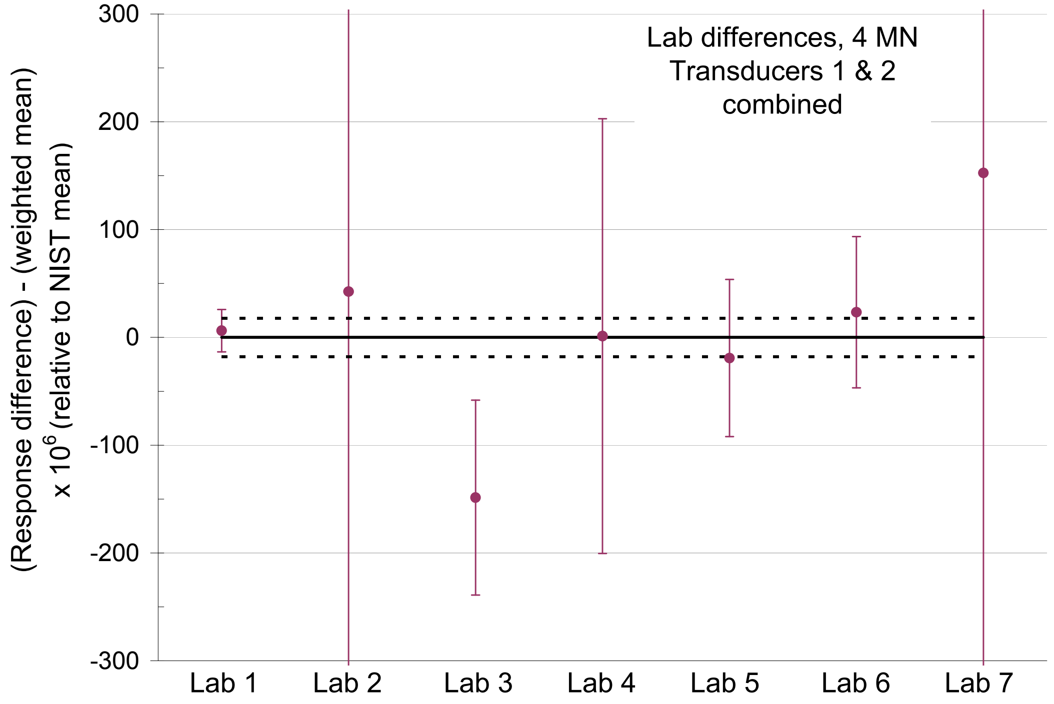

Figure 30. Differences between participant laboratory responses and the weighted mean KCRV, with k=2 expanded uncertainty intervals, calculated from the combined data from Transducers 1 and 2 at the 4 MN force point

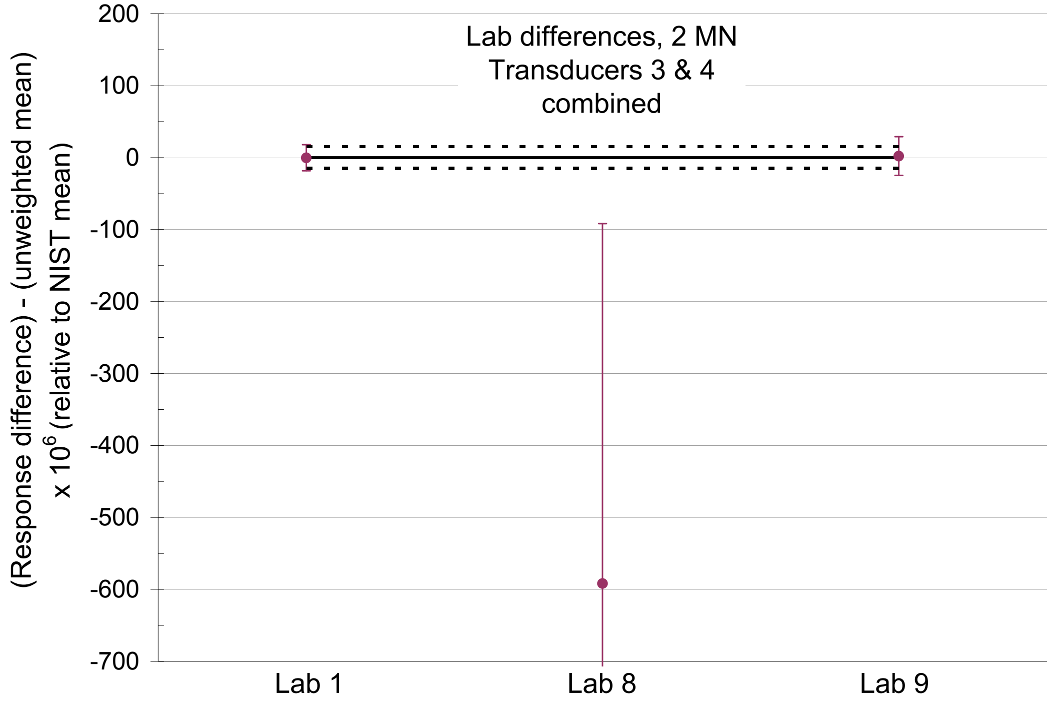

Figure 31. Differences between participant laboratory responses and the weighted mean KCRV, with *k*=2 expanded uncertainty intervals, calculated from the combined data from Transducers 3 and 4 at the 2 MN force point

One more weighted mean calculation from combined data is possible -- for Transducers 1 through 4 for the 2 MN force point. Eqs.(34 & 36) have yielded values of K_k and u_{cKk} for k=1 to 7 from the CCM.F-K4.a data, and for k=1, 8, and 9 from the CCM.F-K4.b data. K_1 is zero for both comparisons, and u_{cK1} is about the same for both. Thus the 2 MN data can be entered into one list for K_k and u_{cKk} for k=1 to 9, and a weighted mean KCRV calculation performed over all nine laboratories by means of Eq.(23-24).

The results of this KCRV calculation at 2 MN from the combined results of Transducers 1 through 4 for all nine participants are shown in Fig. 32.

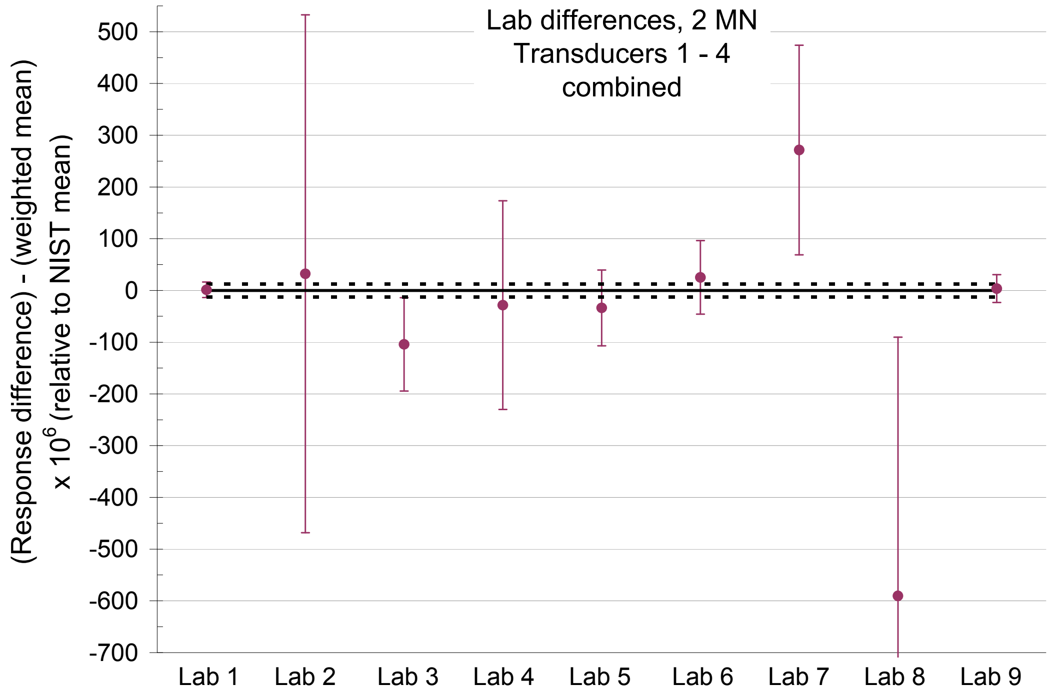

Figure 32. Differences between participant laboratory responses and the weighted mean KCRV, with *k*=2 expanded uncertainty intervals, calculated from the combined data from all four transducers at the 2 MN force point

Equivalence Matrices

The chi-squared consistency test yields probability values of 0.030, 0.055, 0.060, and 0.012 for the analyses shown in Fig. 29, Fig. 30, Fig. 31, and Fig. 32, respectively. Fig. 30 represents the combined results for all seven participating laboratories at the 4 MN force point. Using the criterion that the consistency test fails if the chi-squared probability value is less than 0.05, it is seen that the consistency test passes for CCM.F-K4.a at 4 MN.

The equivalence matrix for the combined results of the seven participants of CCM.F-K4.a is given as Table 14. The values in the table are computed similarly to those in the equivalence matrices of Appendix I, using the seven values of K_k and u_{cKk} that were computed for this combination. Also shown in Table 14 are the differences between each Lab *k* and the 4 MN KCRV.

Table 14. Equivalence Matrix for CCM.F-K4.a, 4 MN, Transducers 1 and 2

4 MN Transducers **1 &** **2**	Δ_{kj} **x 10^6** (relative to NIST mean)						$\mathbf{\Delta}_R$ (LAB - KCRV)	std.unc. in $\mathbf{\Delta}_R$
	Lab 2	Lab 3	Lab 4	Lab 5	Lab 6	Lab 7		
Lab 1	36	-155	-5	-25	17	146	6	13
Lab 2		-191	-41	-62	-19	110	43	250
Lab 3			150	130	172	301	-149	46
Lab 4				-20	22	151	1	101
Lab 5					42	172	-19	38
Lab 6						129	23	36
Lab 7							153	251

4 MN Transducers **1 &** **2**	Standard uncertainty in Δ_{kj} **x 10^6** (relative to NIST mean)					
	Lab 2	Lab 3	Lab 4	Lab 5	Lab 6	Lab 7
Lab 1	250	46	101	38	36	251
Lab 2		254	270	253	253	354
Lab 3			111	58	57	254
Lab 4				107	107	270
Lab 5					51	253
Lab 6						253
Lab 7						

Fig. 32 represents the combined results for all nine participating laboratories at the 2 MN force point. The chi-squared consistency test, yielding a probability value of 0.012, fails for this comparison. However, if Lab 7 is excluded from the KCRV calculation (but not from the comparison with the KCRV), the chi-squared probability value becomes 0.087. Under this condition, the consistency test passes for the combined results at 2 MN. No other statistical tests have been performed to identify outliers.

The equivalence matrix for the combined results of the nine participants, at 2 MN, for both CCM.F-K4.a and CCM.F-K4.b is given as Table 15, along with the differences between each Lab *k* and the 2 MN KCRV.

Table 15. Equivalence Matrix for CCM.F-K4.a & CCM.F-K4.b, 2 MN, Transducers 1-4

2 MN Transducers **1** to **4**	Δ_{kj} **x 10^6** (relative to NIST mean)								Δ_R (LAB - KCRV)	std.unc. in Δ_R
	Lab 2	Lab 3	Lab 4	Lab 5	Lab 6	Lab 7	Lab 8	Lab 9		
Lab 1	31	-105	-30	-35	24	270	-592	2	1	10
Lab 2		-136	-61	-66	-7	239	-623	-28	32	250
Lab 3			76	71	130	376	-486	108	-104	46
Lab 4				-5	54	300	-562	32	-28	101
Lab 5					59	305	-557	37	-34	37
Lab 6						246	-616	-22	25	36
Lab 7							-862	-268	271	101
Lab 8								594	-590	250
Lab 9									4	15

2 MN Transducers **1** to **4**	Standard uncertainty in Δ_{kj} **x 10^6** (relative to NIST mean)							
	Lab 2	Lab 3	Lab 4	Lab 5	Lab 6	Lab 7	Lab 8	Lab 9
Lab 1	250	46	101	37	36	102	250	15
Lab 2		254	270	253	253	270	354	251
Lab 3			110	58	57	111	254	47
Lab 4				107	107	143	270	102
Lab 5					51	108	253	39
Lab 6						107	253	38
Lab 7							270	102
Lab 8								250
Lab 9								

Conclusions from Approach 2

1. The comparison of the seven laboratories capable of achieving the 4 MN force point, using the combined data for Transducers 1 and 2, is depicted in graphical form in Fig.30, with the equivalence matrix given in Table 14. Of the seven participating laboratories, Lab 3 has an expanded uncertainty interval that lies outside of the expanded uncertainty band for the weighted mean KCRV. The chi-squared consistency test passes for this comparison.

2. The comparison of the nine laboratories that conducted measurements at the 2 MN force point, using the combined data for Transducers 1 through 4, is depicted in graphical form in Fig.32, with the equivalence matrix given in Table 15. Two of the nine laboratories, Lab 7 and Lab 8, have expanded uncertainty intervals that lie outside of the expanded uncertainty band for the weighted mean KCRV. The chi-squared consistency test passes for this comparison only if one of the nine laboratories, Lab 7, is excluded from the KCRV calculation.

3. As seen in Fig.30 for the 4 MN force point, the NIST value differs from the weighted mean KCRV by 0.0006 %, which is less than ¼ of the expanded uncertainty associated with this difference. Thus the NIST value, obtained with the only

deadweight machine among the seven participants, could serve as the KCRV without significant change in the conclusions of the comparison.

4. Similarly, as seen in Fig.32 for the 2 MN force point, the NIST value differs from the weighted mean KCRV by 0.0001 %, which is again small relative to the associated expanded uncertainty. The value obtained for the other deadweight machine employed in the 2 MN comparison, Lab 9, is essentially the same as the NIST value, with a difference of 0.0002 %.

5. Analysis Approach 1 and Analysis Approach 2 can only be compared for individual transducers and individual force points. Thus, comparing Fig.17 with Fig.23, for the 2 MN force point for Transducer 1, Approach 1 shows the expanded uncertainty intervals for Lab 3 and Lab 7 as lying outside the expanded uncertainty band for the (unweighted) mean of means KCRV, and Approach 2 shows the expanded uncertainty interval for Lab 7 as lying outside the expanded uncertainty band for the weighted mean KCRV. Comparing Fig.18 with Fig.24, for the 4 MN force point for Transducer 1, Approach 1 also shows the uncertainty intervals for Lab 3 and Lab 7 as lying outside the uncertainty band for the mean of means KCRV, whereas Approach 2 shows the interval for Lab 3 as lying outside the uncertainty band for the weighted mean KCRV. For Transducer 2, comparing Fig.19 with Fig.25 and Fig.20 with Fig.26, Approach 1 shows the expanded uncertainty interval for Lab 6 as lying outside the expanded uncertainty band for the mean of means KCRV for both the 2 MN and 4 MN force points, and Approach 2 shows the expanded uncertainty interval for Lab 6 as lying outside the expanded uncertainty band for the weighted mean KCRV for the 2 MN force point only. Thus there are reasonable correlations between the conclusions of the two analysis approaches.

Acknowledgements

The author is indebted to James Filliben of the Statistical Engineering Division, Information Technology Laboratory at NIST, for his guidance on the statistical analysis of the comparison data presented in this work.

The author is also indebted to Kevin Chesnutwood of the Mechanical Metrology Division, Physical Measurement Laboratory at NIST, for his assistance with conducting the multiple measurements for the key comparison transducers at NIST.

References

[1] Minutes of the meeting of the CCM Force Working Group, held at CSIRO-NML, Lindfield, Sydney, Australia, October 20-23, 1998.

[2] Minutes of the meeting of the CCM Force Working Group, held at BNM-LNE, Paris, France, May 17-18, 1999.

[3] T. Bartel, Uncertainty in NIST Force Measurements, J. Res. Natl. Inst. Stand. Technol. **110** (6) (2005) 589-605.

[4] N. Heckert, J. Filliben, NIST Handbook 148: DATAPLOT Reference Manual, Volume I: Commands, Natl. Inst. Stand. Technol. Handbook Series, 2003.

[5] R. Paule, J. Mandel, Consensus Values and Weighting Factors, J. Res. Natl. Bur. Stand. **87**, (1982) 377-385.

[6] A. Ruhkin, M. Vangel, Estimation of a Common Mean and Weighted Means Statistics, J. American Statistical Association **93** (441) (1998).

[7] S. Schiller, K. Eberhardt, Combining Data from Independent Analysis Methods, Spectrochimica, ACTA **46** (12) (1991).

[8] Graybill, Deal, Combining Unbiased Estimators, Biometrics **15**, (1959) 543-550.

[9] DerSimonain, Laird, Meta-analysis in Clinical Trials, Controlled Clinical Trials **7** (1986) 177-188.

[10] M. Cox, The evaluation of key comparison data, Metrologia **39** (2002), 589-595.