IMEKO 20th TC3, 3rd TC16 and 1st TC22 International Conference
Cultivating metrological knowledge
27th to 30th November, 2007. Merida, Mexico.

Statistical Analysis of 2 MN and 4 MN Force Range Key Comparisons

T. Bartel

National Institute of Standards and Technology (NIST), USA
Telephone: +1-301-975-6461; Fax: +1-301-948-6474; email: twbartel@nist.gov
Postal address: 100 Bureau Drive, Mail Stop 8222, Gaithersburg, MD 20899-8222 USA

Abstract

The factors contributing to the uncertainty are discussed for the measurements that were conducted for the 2 MN and 4 MN force values of the very high force key comparison administered by the Comité International des Poids et Mesures (CIPM). Details of the statistical analysis by the pilot institute, NIST, are provided. Various options for calculating the reference values and degrees of equivalence are presented, accounting for the uncertainty differences associated with the comparison force transfer standards and with the force standard machines of the participating National Metrology Institutes (NMI).

Keywords: Degrees of equivalence, Force, Force standard, Key comparison, Metrology, Reference value, Transfer standard, Uncertainty

1. Introduction

This paper discusses certain aspects of the key comparisons in force designated by the comparison numbers CCM.F-K4.a and CCM.F-K4.b, which correspond to the force values of 4 MN and 2 MN, respectively. This pair of key comparisons is one part of a set of four similar pairs of key comparisons that were initiated by the CIPM through the Consultative Committee for Mass and Related Quantities (CCM). These CIPM key comparisons are being conducted in support of the Mutual Recognition Arrangement (MRA) that was first signed by Member States of the Metre Convention in 1999.

The force ranges for the other CIPM key comparisons in force are: 5 kN and 10 kN, 50 kN and 100 kN, and 0.5 MN and 1 MN. These comparisons have been in progress since 2000, with the measurements for CCM.F-K4.a and CCM.F-K4.b taking place in the time frame from 2002 to 2005.

2. Comparison protocol

The design protocol for the force key comparisons was developed at a meeting of the CCM Force Working Group 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]. NIST was chosen to be the pilot institute for the 2 MN and 4 MN force range comparisons. The protocol called for the pilot institute to carry out the comparisons by circulating a pair of force transducers to each of the other 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.

The participating institutes for CCM.F-K4.a and CCM.F-K4.b selected by the working group are listed in Table 1, along with the types of force standard machine used and their capacities.

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)	Bureau National de Métrologie - Laboratoire National d'Essais (BNM-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

For comparison CCM.F-K4.a, two force transducers were circulated in accordance with the design protocol. These are denoted as Transducer 1 and Transducer 2 in the remainder of this paper. The capacities of these two transducers were 5 MN and 4 MN, respectively. For comparison CCM.F-K4.b, which was begun after the circulation for the 4 MN force range was completed, a second pair of force transducers were circulated, which are denoted as Transducer 3 and Transducer 4 in the remainder of this paper. The capacities of these two transducers were 2 MN and 2.22 MN, respectively.

In addition to the force transducers, a bridge calibration unit was circulated along with the transducers 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, which was specified by the design protocol along with the make and model of the single bridge calibration unit that was circulated.

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. All force points were spaced at six-minute intervals, with the exception of 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.

3. 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. Separate analyses were conducted at each force point for the two transducers employed for each comparison.

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. Separate values are calculated for each of the two transducers circulated to each participant.

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

$$u_d = [s^2/n]^{1/2} \quad . \tag{3}$$

This "data-based" standard uncertainty u_d 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.

Other sources of uncertainty include the standard uncertainty in the applied force, denoted by u_f, and the standard uncertainty in determining the measuring amplifier corrections, denoted by u_v. 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 [2]. 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 combined standard uncertainty u_c (k=1), incorporating all three standard uncertainties u_d, u_f, and u_v, is calculated for each value of r given by Eq.(1) as

$$u_c = [u_d^2 + u_f^2 + u_v^2]^{1/2} \quad , \tag{4}$$

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 net response and the average of the two net 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 and standard deviation, calculated from Eq.(1) and Eq.(2), for each measurement data set obtained from the participants for all four transducers. These values are given in the units of voltage ratio. Also given in the tables are the standard uncertainties in the applied force, which were provided by each participant.

The participating NMIs are not identified by name in order to maintain confidentiality until Draft B is approved by the CCM Working Group on Force. Thus the NMIs are represented in the tables by an arbitrary lab number, with Lab 1 denoting the pilot laboratory NIST. For CCM.F-K4.b, Lab 2 and Lab 3 are not the same as Lab 2 and Lab 3 for CCM.F-K4.a.

The following questions arise, to be answered through analysis of these data:

(1) Are the laboratories equivalent?
(2) Are the transducers equivalent?
(3) Does the measurement protocol, conducted using the transducers employed, yield sufficiently low uncertainty to discern the apparent differences among the laboratories?

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 2	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 3	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

4. NIST results

The results for the measurements at NIST alone are shown in Fig. 1 through Fig. 6, in order to show the variation associated with NIST repeatability or transducer drift with time. Figure 1 through Fig. 4 are from comparison CCM.F-K4.a, for Transducer 1 and Transducer 2 with force points at 2 MN and 4 MN. Figure 5 and Fig. 6 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. 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

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

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. 1 through Fig. 6. Each left-side (solid line) bar represents the data-based expanded uncertainty $U_{dj} = 2u_{dj}$ for the corresponding data set j, where u_{dj} 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.

The following conclusions can be made from Fig. 1 through Fig. 6:

(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 is on the order of ten times "noisier" than Transducer 1.
(3) The differences among data points for Transducer 1 and Transducer 3 are larger than would be accounted for from the data-based uncertainty alone; there may be some source of variability associated with these transducers that is not completely addressed by the measurement statistics.

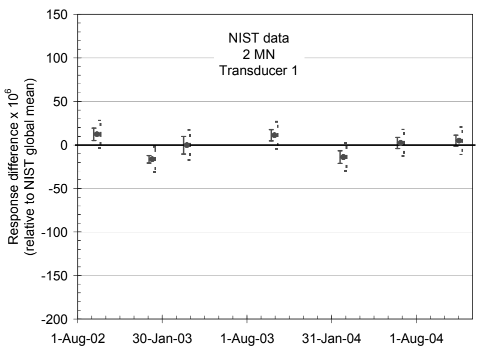

Figure 1. NIST data and k=2 uncertainties at 2 MN force point for Transducer 1

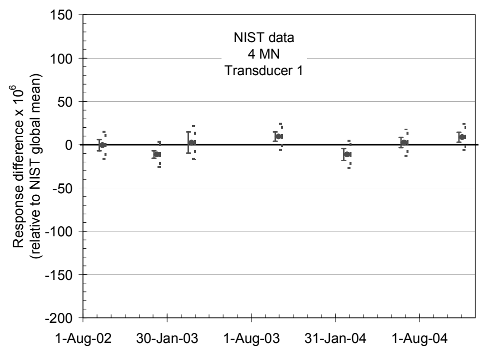

Figure 2. NIST data and k=2 uncertainties at 4 MN force point for Transducer 1

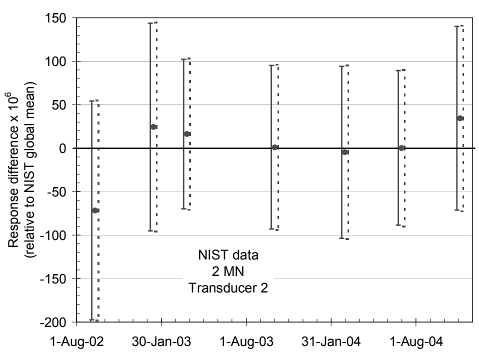

Figure 3. NIST data and k=2 uncertainties at 2 MN force point for Transducer 2

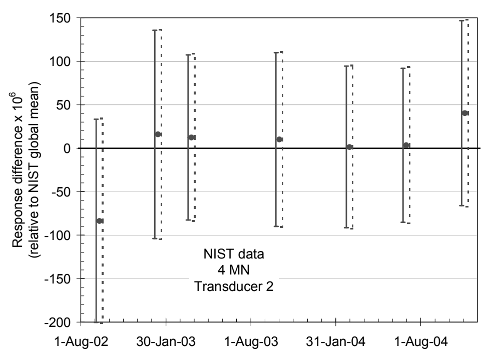

Figure 4. NIST data and k=2 uncertainties at 4 MN force point for Transducer 2

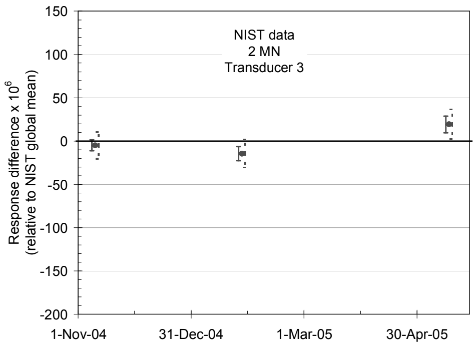

Figure 5. NIST data and k=2 uncertainties at 2 MN force point for Transducer 3

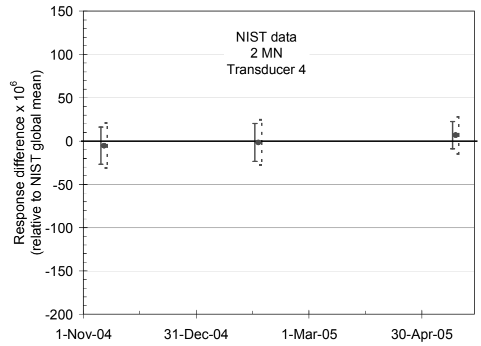

Figure 6. NIST data and k=2 uncertainties at 2 MN force point for Transducer 4

5. Comparison of participating laboratories

A graphical representation of the comparisons of all of the participating laboratories is given in Fig. 7 through Fig. 12. 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. 7 through Fig. 12 are similar to those of Fig. 1 through Fig. 6, 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 is done in order to realize the intention of the "star" circulation for the transducers that was chosen by the CCM Force Working Group. 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

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

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. 7 through Fig. 12, 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. 1 through Fig. 6, two expanded uncertainty intervals are shown for each point in Fig. 7 through Fig. 12. For $k \neq 1$, each left-side (solid line) bar represents the data-based expanded uncertainty $U_{dk} = 2u_{dk}$ for the corresponding data set k, where u_{dk} 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_{d1} and U_{t1} for the pilot laboratory that were most comparable to U_{dk} and U_{tk} for the other laboratories. Thus U_{d1} for each of these figures is taken to be the average data-based expanded uncertainty for the NIST data sets making up the comparison: $U_{d1} = (2/m)\sum u_{dj}$, where the index j represents only the NIST measurement sets, u_{dj} 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 following conclusions can be made from Fig. 7 through Fig. 12:

(1) Transducer 1, Transducer 3, and Transducer 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. 7, Fig. 8, Fig. 11, and Fig. 12, Lab 3 of CCM.F-K4.a and Lab 2 of CCM.F-K4.b are significantly below the pilot Lab 1, and possibly below other estimates of a key comparison reference value, by an amount that is not accounted for by known sources of uncertainty. Lab 7 is correspondingly high, especially at the 2 MN force point.
(3) There are some anomalies, such as the differences in the results at Lab 6 for Transducer 1 and Transducer 2 as shown in Fig. 7 through Fig. 10, and in the relative differences between Lab 1 and Lab 3 for both Transducer 3 and Transducer 4 as seen in Fig. 11 and Fig. 12, that may indicate a transducer-related variability that is not accounted for.

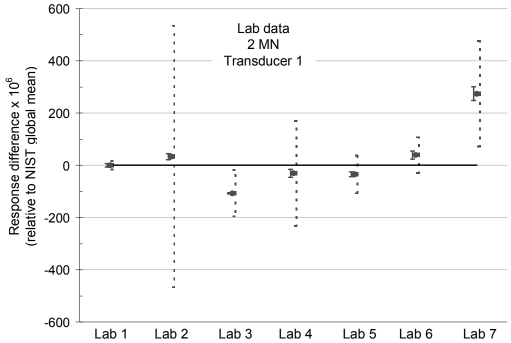

Figure 7. CCM.F-K4.a data and k=2 uncertainties at 2 MN force point for Transducer 1

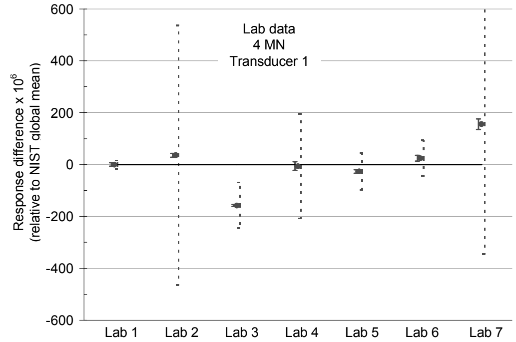

Figure 8. CCM.F-K4.a data and k=2 uncertainties at 4 MN force point for Transducer 1

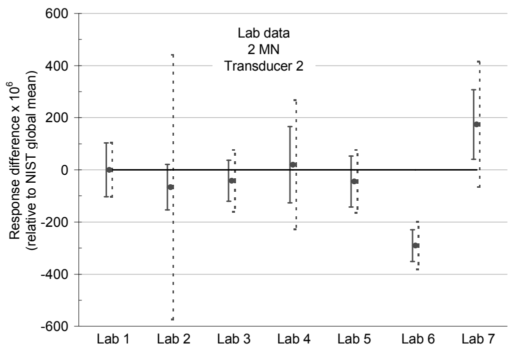

Figure 9. CCM.F-K4.a data and k=2 uncertainties at 2 MN force point for Transducer 2

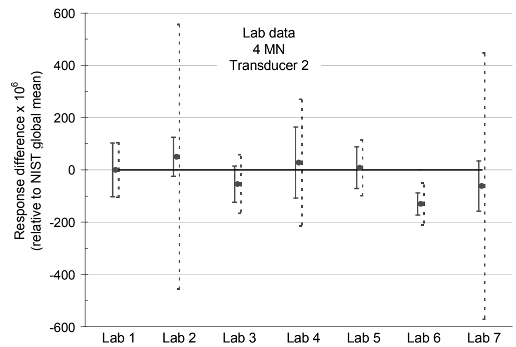

Figure 10. CCM.F-K4.a data and k=2 uncertainties at 4 MN force point for Transducer 2

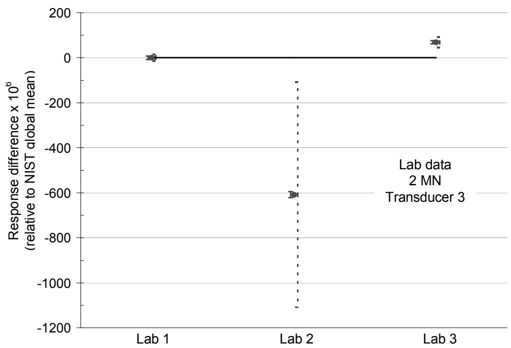

Figure 11. CCM.F-K4.b data and k=2 uncertainties at 2 MN force point for Transducer 3

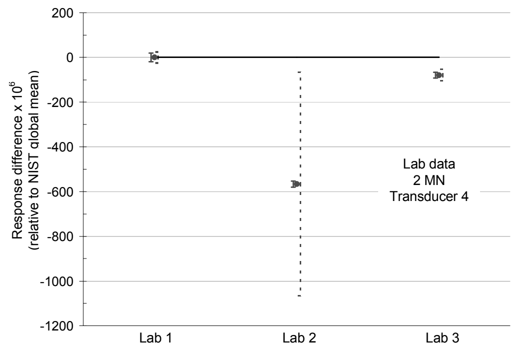

Figure 12. CCM.F-K4.b data and k=2 uncertainties at 2 MN force point for Transducer 4

6. Equivalence matrices

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. A multiplier of 10^6 is used to adjust the scale to presentable values.

$$\Delta_{kj} = (10^6)[d_k - d_j] \ , \qquad (8)$$

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

$$s_{\Delta kj} = (10^6)(s_k^2/n_k + s_j^2/n_j)^{1/2} \ , \qquad (9)$$

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.

For $j = 1$, indicating the "Lab 1" row (for the pilot lab NIST) in the tables, the 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, 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.

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

2 MN Transducer **1**	Δ_{kj} (in indicator units -- mV/V) **x 10⁶**						Standard deviation in Δ_{kj} (in indicator units -- mV/V) **x 10⁶** [calculated from measurement set standard deviations only]					
	Lab 2	Lab 3	Lab 4	Lab 5	Lab 6	Lab 7	Lab 2	Lab 3	Lab 4	Lab 5	Lab 6	Lab 7
Lab 1	26	-86	-25	-28	32	219	5	3	7	5	6	11
Lab 2		-112	-51	-54	5	193		5	8	6	8	11
Lab 3			61	58	117	305			6	4	6	10
Lab 4				-3	56	244				7	9	12
Lab 5					59	247					7	11
Lab 6						187						12
Lab 7												

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

4 MN Transducer **1**	Δ_{kj} (in indicator units -- mV/V) **x 10⁶**						Standard deviation in Δ_{kj} (in indicator units -- mV/V) **x 10⁶** [calculated from measurement set standard deviations only]					
	Lab 2	Lab 3	Lab 4	Lab 5	Lab 6	Lab 7	Lab 2	Lab 3	Lab 4	Lab 5	Lab 6	Lab 7
Lab 1	58	-252	-9	-42	39	249	7	6	15	7	10	16
Lab 2		-309	-67	-100	-19	191		7	15	8	11	17
Lab 3			242	210	291	501			14	6	9	16
Lab 4				-33	48	258				15	16	21
Lab 5					81	291					10	17
Lab 6						210						18
Lab 7												

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

2 MN Transducer **2**	Δ_{kj} (in indicator units -- mV/V) **x 10⁶**						Standard deviation in Δ_{kj} (in indicator units -- mV/V) **x 10⁶** [calculated from measurement set standard deviations only]					
	Lab 2	Lab 3	Lab 4	Lab 5	Lab 6	Lab 7	Lab 2	Lab 3	Lab 4	Lab 5	Lab 6	Lab 7
Lab 1	-66	-42	20	-45	-290	174	62	53	79	59	45	75
Lab 2		24	86	21	-224	240		59	85	65	53	80
Lab 3			62	-3	-248	216			83	63	50	77
Lab 4				-64	-310	154				88	79	99
Lab 5					-246	219					58	83
Lab 6						464						73
Lab 7												

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

4 MN Transducer **2**	Δ_{kj} (in indicator units -- mV/V) **x 10^6**						Standard deviation in Δ_{kj} (in indicator units -- mV/V) **x 10^6** [calculated from measurement set standard deviations only]					
	Lab 2	Lab 3	Lab 4	Lab 5	Lab 6	Lab 7	Lab 2	Lab 3	Lab 4	Lab 5	Lab 6	Lab 7
Lab 1	101	-108	56	16	-260	-124	113	101	151	103	76	118
Lab 2		-209	-44	-84	-361	-224		102	155	109	86	122
Lab 3			165	125	-152	-15			152	105	81	118
Lab 4				-40	-317	-180				157	142	166
Lab 5					-276	-140					90	124
Lab 6						136						105
Lab 7												

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

2 MN Transducer **3**	Δ_{kj} (mV/V) **x 10^6**		Std. dev. in Δ_{kj} (mV/V) **x 10^6**	
	Lab 2	Lab 3	Lab 2	Lab 3
Lab 1	-1207	136	15	12
Lab 2		1343		16
Lab 3				

2 MN Transducer **4**	Δ_{kj} (mV/V) **x 10^6**		Std. dev. in Δ_{kj} (mV/V) **x 10^6**	
	Lab 2	Lab 3	Lab 2	Lab 3
Lab 1	-1022	-144	19	17
Lab 2		878		17
Lab 3				

Note that for Transducer 1 about 90 % of the entries for Δ_{kj} in Table 5 and Table 6 exceed their corresponding 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 the differences among the laboratories. The same conclusion can be made for Transducer 3 and Transducer 4 from the entries in Table 9. 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.

7. Key Comparison Reference Values (KCRV)

7.1 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. 7 through Fig. 12.

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 [3]. 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 [4-8]. 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 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.

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 Schiller-Eberhardt, 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.

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

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

7.2 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 $(1/m)\sum d_k$, 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, the weighted mean is calculated as

$$W = \sum(d_k/u_k^2)/\sum(1/u_k^2). \qquad (11)$$

Two values of the weighted mean were calculated, corresponding to the "data-based" uncertainties and the "total" uncertainties represented by the two sets of uncertainty intervals depicted in Fig. 7 through Fig. 12. For the weighted mean corresponding to the "data-based" uncertainties, the u_k in Eq.(11) are calculated from Eq.(3), using comparison measurement data only. For the weighted mean corresponding to the "total" uncertainties, the u_k in Eq.(11) are the combined standard uncertainties calculated from Eq.(4), incorporating the additional uncertainties associated with the applied forces and the measuring amplifier corrections as discussed in Sec.3.

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.

Since the values in Table 13 are given in the same relative units as plotted in Fig. 7 through Fig. 12, each entry in Table 13 can be considered to represent the ordinate for the corresponding baseline for that key comparison reference value for the figure appropriate to that force-transducer combination.

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.

8. Conclusions

Based on the entire prior analysis, the following are the primary conclusions that may be drawn:

(1) Transducer 1 and Transducer 2 are not equivalent, because of the much larger standard deviations in the data sets for Transducer 2. It may not be valid to combine the results for the comparisons conducted with these transducers.
(2) Transducer 3 and Transducer 4 may be equivalent; however, the population is very small and attempts were not made to combine the results for the corresponding comparisons.
(3) The laboratories are not equivalent (i.e., the differences between laboratories are generally statistically significant) for the comparisons conducted with Transducer 1, Transducer 3, and Transducer 4.

(4) With the possible exception of Lab 6, the laboratories are equivalent for the comparisons conducted with Transducer 2.
(5) No significant drift with time is seen in the transducer characteristics.
(6) On the basis of Transducer 1 for CCM.F-K4.a, Lab 3 is significantly low and Lab 7 is high.
(7) On the basis of Transducer 3 and Transducer 4 for CCM.F-K4.b, Lab 2 is significantly low.
(8) If a key comparison reference value is desired, the unweighted mean may be the most appropriate, with the uncertainty yielded by the consensus mean analysis for the mean of means.

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 Manufacturing Metrology Division, Manufacturing Engineering Laboratory at NIST, for his assistance with the acquisition of the data from the multiple measurements conducted with 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] T. Bartel, Uncertainty in NIST Force Measurements, J. Res. Natl. Inst. Stand. Technol. **110** (6) (2005) 589-603.

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

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

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

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

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

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