Metrologia **46** (2009) 345–350 doi:10.1088/0026-1394/46/3/024

A generalized method for the multiple artefacts problem in interlaboratory comparisons with linear trends

Weiping Zhang[1,2], **Nien Fan Zhang**[1] **and Hung-kung Liu**[1]

[1] Statistical Engineering Division, National Institute of Standards and Technology, Gaithersburg, MD 20899, USA
[2] Department of Statistics and Finance, University of Science and Technology of China, Hefei, 230026, People's Republic of China

Received 25 November 2008, in final form 19 April 2009
Published 22 May 2009
Online at stacks.iop.org/Met/46/345

Abstract
A generalized statistical approach for interlaboratory comparisons with linear trends is proposed. This new approach can be applied to the general case when the artefacts are measured and reported multiple times in each participating laboratory. The advantages of this approach are that it is consistent with the previous approaches when only the pilot lababoratory makes multiple measurements and it applies whether or not there exists a trend. The uncertainties for the comparison reference value and the degree of equivalence are also provided. As an illustration, the method is applied to the SIM.EM-K2 comparison for resistance at the level of 1 GΩ.

S This article has associated online supplementary data files

1. Introduction

Interlaboratory comparisons are conducted for intercomparison of measurement results among laboratories. Key comparisons are special interlaboratory comparisons that serve as technical bases for the Mutual Recognition Arrangement (MRA) of the International Committee for Weights and Measures (CIPM) between national metrology institutes (NMIs) and regional metrology organizations (RMOs) around the world [1]. In some key comparisons, the measurand has a trend or a drift and thus the measurements of the transport artefacts made by the participating NMIs will show the trend. Zhang *et al* proposed an approach for the analysis of single [2] and multiple artefact problems [3] with linear trends. Elster *et al* [4] proposed a procedure for the analysis of key comparison data in the case where the travelling artefact shows a deterministic or random drift. Stepanov [5] compares three approaches including the one in [2] for evaluating key comparisons with linear trends. In the scenario discussed in [2, 3], a measurand was measured by the pilot NMI in several separate time periods and at only one time by the other participating NMIs. Since the measured values show a linear trend in time, a simple linear model with the same slope and different intercepts for each NMI was proposed. Only the measured values of the pilot NMI were used to estimate the joint slope as in [6] for CCEM.EM-K2.

Along with key comparisons among the NMIs, many RMOs are conducting interlaboratory comparisons between corresponding laboratories in their regions. In some RMO comparisons, some non-pilot laboratories also measure an artefact multiple times. For example, recently the Technical Committee of the Inter-American Metrology System (SIM) for Electricity and Magnetism conducted a resistance intercomparison at the 1 Ω, 1 MΩ and 1 GΩ levels between six SIM laboratories [7, 8], which are the National Institute of Standards and Technology (NIST) of the United States, Instituto Nacional de Technologia Industrial (INTI) of Argentina, National Institute of Metrology Standardization and Industrial Quality (INMETRO) of Brazil, Administración Nacional de Usinas y Transmisiones Electricas (UTE) of Uruguay, National Research Council (NRC) of Canada, and Centro Nacional de Metrologia (CENAM) of Mexico. In the comparison, the pilot laboratory, NIST, made measurements in five separate periods. In SIM.EM-K2 for the 1 GΩ comparison, all other non-pilot NMIs made measurements in two separate time periods (about six months apart) except UTE. The SIM Working Group made the decision to allow each participant to measure the travelling standards twice during the comparison. As reported in [8], this provided additional information about the linear drift of the standard resistors in addition to information from just the pilot NMI.

In this case, the approach proposed in [2, 3] cannot be applied without sacrificing some information and thus a more general methodology is needed. In this paper, we propose a general approach to deal with this case. The methodology is consistent with the approaches in [2, 3] when the pilot laboratory is the only one making measurements in multiple time periods. It is also consistent with the method for the trivial case when the trend reduces to zero. As in [3], multiple artefacts are also considered.

2. Models and parameter estimation

In this paper, L travelling artefacts are considered. As in [3], we assume that for a fixed artefact, the measurements of any particular laboratory have a linear trend in time, and the slopes of the trends for all p laboratories are the same, while we allow different intercepts for different laboratories. Specifically, we assume that for all artefacts the ith laboratory ($i = 1, \ldots, p$) makes k_i measurements with $k_i \geqslant 1$. For the lth artefact, the jth measurement (or the average of the measurements) made at laboratory i, $X_{ij}(l)$ is measured at the time $t_{ij}(l)$ ($j = 1, \ldots, k_i$). Without loss of generality, we assume that the pilot laboratory is the first one among all p laboratories with $k_1 > 1$.

For a fixed artefact, say l ($l = 1, \ldots, L$), we assume a simple linear regression holds for all the measurements, i.e.

$$X_{ij}(l) = \alpha_i(l) + \beta(l)t_{ij}(l) + e_{ij}(l), \tag{1}$$

for $j = 1, \ldots, k_i$, $i = 1, \ldots, p$ and $l = 1, \ldots, L$. For a fixed artefact, say the lth artefact, we further assume that for each laboratory, say the ith laboratory, an error of the measurement $X_{ij}(l)$ can be expressed as

$$e_{ij}(l) = e_{ij,A}(l) + (1 - I_i(l))e_{ij,B}(l) + I_i(l)e_{i,B}(l). \tag{2}$$

In (2), the indicator $I_i(l) = 1$ when the errors $e_{ij,B}(l)$ are the same for all the measurements made by the ith laboratory and $I_i(l) = 0$ otherwise. The error components $e_{ij,A}(l)$ and $(e_{i,B}(l), e_{ij,B}(l))$ are statistically independent of each other with standard uncertainties of $\sigma_{ij,A}(l)$ and $(\sigma_{i,B}(l), \sigma_{ij,B}(l))$, which are type A and type B evaluations of standard uncertainty, respectively. As discussed in [2], $\sigma_{ij,A}(l)$ also denotes the standard deviation. By the same token, we do not distinguish the notation of a variance and the square of a standard uncertainty in this paper. Equation (2) indicates that the measurements of different artefacts (whether by the same or by different laboratories) are statistically independent, while the measurements for the same artefact, made at the same laboratory can be independent or not, depending on the indicator $I_i(l)$. From (2), when $I_i(l) = 1$, type B uncertainties are the same for all the measurements made by the ith laboratory. In contrast, when $I_i(l) = 0$, type B uncertainties may be not the same for all the measurements made by the ith laboratory.

Using matrix notation, (1) can be expressed as

$$X(l) = Z(l)\theta(l) + \varepsilon(l), \tag{3}$$

where

$$X(l) = (X_{11}(l), \ldots, X_{1k_1}(l); \ldots; X_{p1}(l), \ldots, X_{pk_p}(l))', \tag{4}$$

$$\theta = (\alpha_1(l), \ldots, \alpha_p(l), \beta(l))', \tag{5}$$

$$Z(l) = \begin{pmatrix} 1 & 0 & \cdots & 0 & t_{11}(l) \\ & & & & \vdots \\ 1 & 0 & \cdots & 0 & t_{1k_1}(l) \\ 0 & 1 & \cdots & 0 & t_{21}(l) \\ & & \vdots & & \vdots \\ 0 & 1 & \cdots & 0 & t_{2k_2}(l) \\ & & \vdots & & \\ 0 & \cdots & 0 & 1 & t_{p1}(l) \\ & & & & \vdots \\ 0 & \cdots & 0 & 1 & t_{pk_p}(l) \end{pmatrix} \tag{6}$$

is a $(k_1 + k_2 + \cdots + k_p)$ by $p + 1$ matrix, and

$$\begin{aligned} \varepsilon(l) = {} & (e_{11,A}(l) + I_1(l)e_{1,B}(l) + (1 - I_1(l))e_{11,B}(l), \ldots, \\ & e_{1k_1,A}(l) + I_1(l)e_{1,B}(l) + (1 - I_1(l)) \\ & \times e_{1k_1,B}(l); \ldots; e_{p1,A}(l) + I_p(l)e_{p,B}(l) + (1 - I_p(l)) \\ & \times e_{p1,B}(l), \ldots, e_{pk_p,A}(l) + I_p(l)e_{p,B}(l) \\ & + (1 - I_p(l))e_{pk_p,B}(l))' \end{aligned} \tag{7}$$

with mean $E[\varepsilon(l)] = 0$ and the covariance matrix

$$\begin{aligned} \mathrm{Cov}[\varepsilon(l)] \triangleq \Sigma(l) = {} & \mathrm{diag}\,\{\mathrm{diag}\{\sigma_{11,A}^2(l) + (1 - I_1(l)) \\ & \times\sigma_{11,B}^2(l), \ldots, \sigma_{1k_1,A}^2(l) + (1 - I_1(l))\sigma_{1k_1,B}^2(l)\} \\ & + I_1(l)\sigma_{1,B}^2(l)J_{k_1}; \ldots; \mathrm{diag}\{\sigma_{p1,A}^2(l) + (1 - I_p(l)) \\ & \times \sigma_{p1,B}^2(l), \ldots, \sigma_{pk_p,A}^2(l) + (1 - I_p(l))\sigma_{pk_p,B}^2(l)\} \\ & + I_p(l)\sigma_{p,B}^2(l)J_{k_p}\}. \end{aligned} \tag{8}$$

We use ξ' to denote the transpose of a vector ξ. $J_k = 1_k 1_k'$ with $1_k = (1, \ldots, 1)'$ of length k. The matrix $\mathrm{diag}\{c_1, \ldots, c_n\}$ is a diagonal matrix with elements $c_1, \ldots, c_n$. In appendix 1 (appendices 1–3 are available from the electronic verison of this journal), we re-express $X(l)$, $\varepsilon(l)$ and $\mathrm{Cov}[\varepsilon(l)]$ and make them more explicit. It is obvious that for a fixed laboratory, say i, and a fixed artefact, say l, when $I_i(l) = 1$, then its k_i measurements $\{X_{ij}(l); j = 1, \ldots, k_i\}$ are not independent of each other since from (2) they have the same random error $e_{i,B}(l)$; while when $I_i(l) = 0$, these k_i times measurements $\{X_{ij}(l); j = 1, \ldots, k_i\}$ are independent of each other.

It is well known as in [9], p 230 that the best linear unbiased estimator of $\theta(l)$ in (3) is the generalized least square estimator, i.e.

$$\hat{\theta}(l) = (Z(l)'\Sigma^{-1}(l)Z(l))^{-1}Z(l)'\Sigma^{-1}(l)X(l). \tag{9}$$

After laborious but straightforward mathematical operations sketched in appendices 1 and 2, the estimators of $\alpha_i(l)$ ($i = 1, \ldots, p$) and $\beta(l)$ can be written as

$$\hat{\alpha}_i(l) = X_i(l) - \hat{\beta}(l)t_i(l), \qquad i = 1, \ldots, p \tag{10}$$

$$\hat{\beta}(l) = \frac{\sum_{i=1}^{p}\sum_{j=1}^{k_i} \frac{(t_{ij}(l) - t_i(l))(X_{ij}(l) - X_i(l))}{\sigma_{ij,A}^2(l) + (1 - I_i(l))\sigma_{ij,B}^2(l)}}{\sum_{i=1}^{p}\sum_{j=1}^{k_i} \frac{(t_{ij}(l) - t_i(l))^2}{\sigma_{ij,A}^2(l) + (1 - I_i(l))\sigma_{ij,B}^2(l)}}, \tag{11}$$

where

$$t_i(l) = \sum_{j=1}^{k_i} w_{ij}(l) t_{ij}(l), \qquad X_i(l) = \sum_{j=1}^{k_i} w_{ij}(l) X_{ij}(l) \tag{12}$$

are weighted means of $\{t_{ij}(l)\}$ and $\{X_{ij}(l)\}$ with

$$w_{ij}(l) = \frac{1/(\sigma_{ij,A}^2(l) + (1 - I_i(l))\sigma_{ij,B}^2(l))}{\sum_{j=1}^{k_i} 1/(\sigma_{ij,A}^2(l) + (1 - I_i(l))\sigma_{ij,B}^2(l))},$$

respectively. From (2) and (12), the corresponding uncertainty for $X_i(l)$, $u_i(l)$, for the lth artefact in the ith laboratory is given by

$$u_i^2(l) = \frac{1}{\sum_{j=1}^{k_i} 1/(\sigma_{ij,A}^2(l) + (1 - I_i(l))\sigma_{ij,B}^2(l))} + I_i(l)\sigma_{i,B}^2(l). \tag{13}$$

The estimators of the linear regressions $\hat{\alpha}_i(l)$ and $\hat{\beta}(l)$ in (10) and (11) are based on the measurements made by all laboratories. The above estimators are consistent with the estimators in [2, 3], where $I_i(l) = 1$, $\sigma_{ij,B}^2(l) = \sigma_{i,B}^2$ for all i, j, l and $k_i = 1$ when $i \neq 1$. However, $\hat{\alpha}_i(l)$ and $\hat{\beta}(l)$ given in [2, 3] are different from those given in the general case here, since as shown in (11), $X_{ij}(l)$, $\sigma_{ij,A}^2(l)$ and $\sigma_{ij,B}^2(l)$ for each laboratory are needed to estimate the regression parameters.

Similar to the results in [2, 3], the corresponding uncertainties for the estimators in (10) and (11) are given by

$$u_{\hat{\beta}(l)}^2 = \frac{1}{\sum_{i=1}^{p} \sum_{j=1}^{k_i} \frac{(t_{ij}(l) - t_i(l))^2}{\sigma_{ij,A}^2(l) + (1 - I_i(l))\sigma_{ij,B}^2(l)}}, \tag{14}$$

$$u_{\hat{\alpha}_i(l)}^2 = u_i^2(l) + u_{\hat{\beta}(l)}^2 t_i^2(l), \tag{15}$$

$$\mathrm{Cov}[X_i(l), \hat{\beta}(l)] = 0, \tag{16}$$

$$\mathrm{Cov}[\hat{\alpha}_i(l), \hat{\beta}(l)] = \frac{-t_i(l)}{\sum_{i=1}^{p} \sum_{j=1}^{k_i} \frac{(t_{ij}(l) - t_i(l))^2}{\sigma_{ij,A}^2(l) + (1 - I_i(l))\sigma_{ij,B}^2(l)}} \quad \text{for } i = 1, \ldots, p \tag{17}$$

and

$$\mathrm{Cov}[\hat{\alpha}_i(l), \hat{\alpha}_j(l)] = \frac{t_i(l) t_j(l)}{\sum_{i=1}^{p} \sum_{j=1}^{k_i} \frac{(t_{ij}(l) - t_i(l))^2}{\sigma_{ij,A}^2(l) + (1 - I_i(l))\sigma_{ij,B}^2(l)}} \quad \text{for } i \neq j. \tag{18}$$

The predicted value based on the lth regression line for the value of the lth artefact and the ith laboratory at any time $t(l)$ is given by

$$L_{il}(t(l)) = \hat{\alpha}_i(l) + \hat{\beta}(l) t(l). \tag{19}$$

The corresponding uncertainty is given by

$$u_{\hat{\alpha}_i(l) + \hat{\beta}(l)t}^2 = u_i^2(l) + u_{\hat{\beta}(l)}^2 (t_i(l) - t(l))^2. \tag{20}$$

The derivations of (19) and (20) are given in appendix 3.

3. Comparison reference value

In section 2, a regression line is established for each artefact and each laboratory. As discussed in [2, 3], the comparison reference value as a weighted mean of the predicted values over the artefacts and laboratories is time dependent. For the comparison reference value (CRV) at any time t (denoted by CRV_t), as in [3], we use a weighted mean of $\hat{\alpha} + \hat{\beta}t$ over all the laboratories $i = 1, \ldots, p$ and all the artefacts $l = 1, \ldots, L$, i.e.

$$\mathrm{CRV}_{\vec{t}}(\omega, \nu) = \sum_{i=1}^{p} \omega_i \left(\sum_{l=1}^{L} \nu_{il} L_{il}(t(l)) \right), \tag{21}$$

where the time t is allowed to be different for different artefacts, i.e. $\vec{t} = (t(1), \ldots, t(l), \ldots, t(L))$ and the predicted value $L_{il}(t(l))$ is given by (19). The weights $\omega = (\omega_1, \ldots, \omega_p)$ represent the effects of all the participating laboratories. The weight ν_{il} for the ith laboratory and the lth artefact satisfy $\sum_{l=1}^{L} \nu_{il} = 1$ for each fixed i and also $\sum_{i=1}^{p} \omega_i = 1$ for the participating laboratories. An alternative to (21) is to take a weighted average of the prediction values over laboratories for each artefact first and then take a weighted average over artefacts. As discussed in [3], when there is no trend for a key comparison with multiple artefacts, a comparison reference value is often calculated by first taking a weighted average of the measurements (or the difference between the measurements and the nominal value) for all L artefacts within each laboratory and then taking a weighted average over all laboratories, e.g. CCPR-K2.a [10]. Therefore, we will consider only the first way. Similar to (16) in [3], the variance of CRV at time $\vec{t} = (t(1), \ldots, t(L))$ is given by

$$\begin{aligned}
&\mathrm{Var}[\mathrm{CRV}_{\vec{t}}(\omega, \nu)] \\
&= \mathrm{Var}\left[\sum_{i=1}^{p} \omega_i \left(\sum_{l=1}^{L} \nu_{il} X_i(l) + \hat{\beta}(l)(t(l) - t_i(l)) \right) \right] \\
&= \mathrm{Var}\left[\sum_{i=1}^{p} \omega_i \sum_{l=1}^{L} \nu_{il} X_i(l) \right] \\
&\quad + \sum_{l=1}^{L} \left[\sum_{i=1}^{p} \omega_i \nu_{il} (t(l) - t_i(l)) \right]^2 u_{\hat{\beta}(l)}^2. \\
&= \sum_{i=1}^{p} \omega_i^2 \sum_{l=1}^{L} \nu_{il}^2 u_i^2(l) \\
&\quad + \sum_{l=1}^{L} \left[\sum_{i=1}^{p} \omega_i \nu_{il} (t(l) - t_i(l)) \right]^2 u_{\hat{\beta}(l)}^2.
\end{aligned} \tag{22}$$

In metrology it is commonly assumed that the weights ν_{il} do not depend on the laboratory. Namely, $\nu_{il} = \nu_l$ for $i = 1, \ldots, p$ as in CCEM.EM-K2 [6]. As discussed in [2, 3], the uncertainty given in (22) depends on $\vec{t}$ as well as the weight $\omega = (\omega_1, \ldots, \omega_p)$ and ν_{il}. For a fixed set of $\{\nu_l\}$, we use the criterion of minimizing the variance of CRV in (22) to find the optimal weights $\{\omega_i\}$and the corrsponding $\vec{t}$. The second term in the last equality of (22) will vanish when choosing

$$t(l) = t_\omega(l) = \sum_{i=1}^{p} \omega_i t_i(l) \tag{23}$$

for $l = 1, \ldots, L$. With this choice, from (21), (19) and (10),

$$\mathrm{CRV}_{\vec{t}_\omega}(\omega, \nu) = \sum_{i=1}^{P} \omega_i \sum_{l=1}^{L} \nu_l X_i(l). \tag{24}$$

From (22), the corresponding uncertainty of $\mathrm{CRV}_{\vec{t}_\omega}(\omega, \nu)$ is given by

$$u^2_{\mathrm{CRV}_{\vec{t}_\omega}(\omega,\nu)} = \sum_{i=1}^{p} \omega_i^2 \sum_{l=1}^{L} \nu_l^2 u_i^2(l). \tag{25}$$

As in the no-trend case, for a fixed set of ν_l, $u^2_{\mathrm{CRV}_{\vec{t}_\omega}(\omega,\nu)}$ is minimized when the weights $\{\omega_i\}$ are given by

$$\omega_i^*(\nu) = \frac{1/\sum_{l=1}^{L} \nu_l^2 u_i^2(l)}{\sum_{i=1}^{p}\left[1/\sum_{l=1}^{L} \nu_l^2 u_i^2(l)\right]}. \tag{26}$$

See [11]. With this set of weights $\{\omega_i^*(\nu)\}$, the corresponding $\vec{t}^* = (t^*(1), \ldots, t^*(l), \ldots, t^*(L))$ from (23) is

$$t^*(l) = \sum_{i=1}^{p} \omega_i^*(\nu) t_i(l), \qquad l = 1, \ldots, L \tag{27}$$

and the corresponding CRV in (24) is given by

$$\mathrm{CRV}_{\vec{t}^*}(\omega^*, \nu) = \sum_{i=1}^{P} \omega_i^*(\nu) \sum_{l=1}^{L} \nu_l X_i(l). \tag{28}$$

The standard uncertainty of this CRV is given by

$$u^2_{\mathrm{CRV}_{\vec{t}^*}(\omega^*,\nu)} = \frac{1}{\sum_{i=1}^{p}\left[1/\sum_{l=1}^{L} \nu_l^2 u_i^2(l)\right]}. \tag{29}$$

In practice, a choice of ν_l can be formed by the 'mean-square residuals' for the lth regression line for the pilot laboratory as in CCEM-K2, i.e.

$$\nu_l = \frac{1/\rho^2(l)}{\sum_{l=1}^{L} 1/\rho^2(l)}, \tag{30}$$

where

$$\rho^2(l) = \frac{\sum_{j=1}^{k_1}(x_{1j}(l) - \hat{\alpha}_1(l) - \hat{\beta}(l)t_{1j}(l))^2}{k_1 - 2}. \tag{31}$$

4. Degrees of equivalence

4.1. Degrees of equivalence of the national measurement standards with respect to the CRV

For the degrees of equivalence of the national measurement standards from the ith laboratory with respect to the CRV, we only consider the case when $\nu_{il} = \nu_l$, $\omega_i = \omega_i^*(\nu)$ and $\vec{t} = \vec{t}^*$ as given by (26) and (27). The degree of equivalence of the national measurement standard from the ith laboratory with respect to the $\mathrm{CRV}_{\vec{t}^*}(\omega^*, \nu)$ is defined as the difference

$$D_{i,\mathrm{CRV}_{\vec{t}^*}(\omega^*,\nu)} = \sum_{l=1}^{L} \nu_l(\hat{\alpha}_i(l) + \hat{\beta}(l)t^*(l)) - \mathrm{CRV}_{\vec{t}^*}(\omega^*, \nu). \tag{32}$$

Similar to (31) in [3], from (20), (16)–(18) and (29) the corresponding standard uncertainty is given by

$$\begin{aligned}
\mathrm{Var}[D_{i,\mathrm{CRV}_{\vec{t}^*}(\omega^*,\nu)}] &= \sum_{l=1}^{L} \nu_l^2 \mathrm{Var}[\hat{\alpha}_i(l) + \hat{\beta}(l)t^*(l)] \\
&\quad + \mathrm{Var}[\mathrm{CRV}_{\vec{t}^*}(\omega^*, \nu)] \\
&\quad - 2\mathrm{Cov}\left[\sum_{l=1}^{L} \nu_l(\hat{\alpha}_i(l) + \hat{\beta}(l)t^*(l)), \mathrm{CRV}_{\vec{t}^*}(\omega^*, \nu)\right] \\
&= \sum_{l=1}^{L} \nu_l^2 \left[u_i^2(l) + \frac{(t_i(l) - t^*(l))^2}{\sum_{i=1}^{p}\sum_{j=1}^{k_i} \frac{(t_{ij}(l)-t_i(l))^2}{\sigma^2_{ij,A}(l)+(1-I_i(l))\sigma^2_{ij,B}(l)}}\right] \\
&\quad + \frac{1}{\sum_{i=1}^{p}\left[1/\sum_{l=1}^{L} \nu_l^2 u_i^2(l)\right]} - 2\sum_{l=1}^{L} \nu_l^2 \omega_i^*(\nu) u_i^2(l) \\
&= (1 - 2\omega_i^*(\nu)) \sum_{l=1}^{L} \nu_l^2 u_i^2(l) \\
&\quad + \sum_{l=1}^{L} \frac{\nu_l^2 (t_i(l) - t^*(l))^2}{\sum_{i=1}^{p}\sum_{j=1}^{k_i} \frac{(t_{ij}(l)-t_i(l))^2}{\sigma^2_{ij,A}(l)+(1-I_i(l))\sigma^2_{ij,B}(l)}} \\
&\quad + \frac{1}{\sum_{i=1}^{p}\left[1/\sum_{l=1}^{L} \nu_l^2 u_i^2(l)\right]}.
\end{aligned} \tag{33}$$

4.2. Degrees of equivalence between pairs of national measurement standards

The degree of equivalence between two national measurement standards at time $\vec{t}$ is defined as in [3], i.e.

$$\begin{aligned}
D_{i,j} &= \sum_{l=1}^{L} \nu_l[(\hat{\alpha}_i(l) + \hat{\beta}(l)t(l)] - \sum_{l=1}^{L} \nu_l[(\hat{\alpha}_j(l) + \hat{\beta}(l)t(l))] \\
&= \sum_{l=1}^{L} \nu_l[\hat{\alpha}_i(l) - \hat{\alpha}_j(l)],
\end{aligned} \tag{34}$$

when $i \neq j$. Thus the quantity is independent of $\vec{t}$. Since $\hat{\alpha}_i(l)$ are independent for different l, by (14), (15) and (18) the corresponding standard uncertainty is given by

$$\begin{aligned}
u^2_{D_{i,j}} &= \sum_{l=1}^{L} \nu_l^2[u^2_{\hat{\alpha}_i(l)} + u^2_{\hat{\alpha}_j(l)}] - 2\sum_{l=1}^{L} \nu_l^2 \mathrm{Cov}[\hat{\alpha}_i(l), \hat{\alpha}_j(l)] \\
&= \sum_{l=1}^{L} \nu_l^2[u_i^2(l) + u_j^2(l)] \\
&\quad + \sum_{l=1}^{L} \nu_l^2 \left[\frac{(t_i(l) - t_j(l))^2}{\sum_{i=1}^{p}\sum_{j=1}^{k_i} \frac{(t_{ij}(l)-t_i(l))^2}{\sigma^2_{ij,A}(l)+(1-I_i(l))\sigma^2_{ij,B}(l)}}\right].
\end{aligned} \tag{35}$$

The corresponding expanded uncertainty is given by two times the standard uncertainty.

5. An example

To illustrate the approach, we applied it to the SIM.EM-K2 comparison for resistance at the level of 1 GΩ in [7]. The Working Group for Electricity and Magnetism of the SIM

Table 1. Information for the standard S/N 9104.

Lab	Mean date of measurement (year)	Measurement result/10^{-6}	Type A standard uncertainty/ 10^{-6}	Type B standard uncertainty/ 10^{-6}
NIST	2005.95	16.53	0.86	2.69
INTI	2006.05	−4.42	8.00	7.32
INMETRO	2006.13	13.10	7.00	6.09
UTE	2006.28	13.20	2.32	22.12
NIST	2006.41	21.34	0.88	2.69
NRC	2006.50	15.60	1.33	12.50
CENAM	2006.72	23.80	1.00	17.58
NIST	2006.82	20.89	1.35	2.69
INTI	2006.92	20.83	0.60	7.35
INMETRO	2007.06	15.00	3.31	6.12
NIST	2007.22	24.08	1.12	2.69
NRC	2007.36	13.90	0.43	10.58
CENAM	2007.53	27.00	0.78	10.09
NIST	2007.62	22.72	0.92	2.69

Table 2. Information for the standard S/N 9105.

Lab	Mean date of measurement (year)	Measurement result/10^{-6}	Type A uncertainty/ 10^{-6}	Type B uncertainty/ 10^{-6}
NIST	2005.95	−22.53	1.39	2.69
INTI	2006.05	−45.35	8.00	9.41
INMETRO	2006.13	−13.80	6.80	6.98
UTE	2006.28	−22.00	1.47	22.12
NIST	2006.41	−17.69	1.65	2.69
NRC	2006.50	−23.60	1.58	12.58
CENAM	2006.72	−9.00	2.00	23.00
NIST	2006.82	−12.48	2.11	2.69
INTI	2006.92	−16.88	0.80	9.39
INMETRO	2007.06	−19.80	3.89	6.51
NIST	2007.22	−13.86	2.11	2.69
NRC	2007.36	−19.60	0.75	10.58
CENAM	2007.53	−11.00	1.40	10.17
NIST	2007.62	−14.73	1.31	2.69

initiated the key and supplemental comparisons SIM.EM-K1-K2-S6 to provide the first internationally recognized comparisons of precision resistance measurements for nations of the western hemisphere. Six NMIs participated in the comparisons. NIST provided the comparison standards and acted as the pilot laboratory. Two travelling standards of NIST designed film-type standard resistors were used. During the comparison, the two transport standards were measured at NIST for five time periods. For each period, an average value of the dates when the measurements were made is calculated and called a mean date of measurement. Each of the five non-pilot laboratories made measurements at two separate time periods except UTE which only measured at one time period. An uncertainty budget that includes type A and type B evaluations of standard uncertainties for each NMI's measurement process was also reported. Tables 1 and 2 list the information for the two travelling 1 GΩ standards S/N 9104 and S/N 9105 in SIM-EM-K2 comparison, respectively. The measurement results are listed as the relative differences between the measured values and the norminal value 1 GΩ in the unit of 1×10^{-6}. The uncertainties are the standard uncertainties, i.e. $k = 1$.

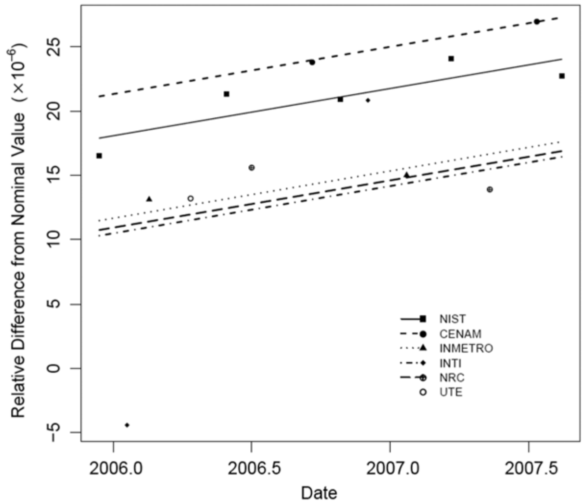

Figure 1. Measurements of 1 GΩ standard S/N 9104 by all participants and the regression lines.

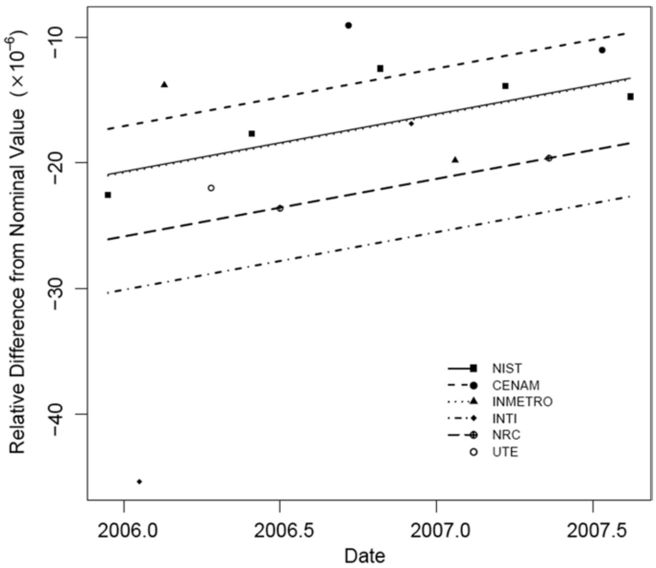

Figure 2. Measurements of 1 GΩ standard S/N 9105 by all participants and the regression lines.

Table 3. The degrees of equivalence of the national measurement standards with respect to the CRV and their expanded uncertainties ($\times 10^{-6}$).

	NIST	INTI	INMETRO	UTE	NRC	CENAM
$D_{i,\mathrm{CRV}_{t*}}$	1.9388	−6.1095	−2.9151	−3.1417	−4.7230	5.2783
$2u_{D_{i,\mathrm{CRV}}}$	2.7190	9.3076	8.2212	35.0568	12.3852	13.5984

Notice that in the SIM.EM-K2 comparison, type B standard uncertainty of the measurements made at different time periods may not be the same for some laboratories as shown in tables 1 and 2. In practice whether an indicator $I_i(l)$ in equation (2) corresponding to the lth artefact and the ith laboratory takes 1 or 0 depends on the setting of the measurement process and thus the uncertainty budget. For the lth artefact, the indicator $I_i(l) = 1$ if the error $e_{ij,B}(l)$ in (2) for measurements made at different time periods are the same for the ith laboratory and $I_i(l) = 0$ otherwise. This leads to type B standard uncertainties for measurements made for the

Table 4. The degrees of equivalence of pairs of national measurement standards with respect to their expanded uncertainties in the parentheses ($\times 10^{-6}$).

	NIST	INTI	INMETRO	UTE	NRC	CENAM
NIST		8.0484 (10.8010)	4.8539 (9.8806)	5.0805 (35.4822)	6.6618 (13.5466)	−3.3395 (14.6630)
INTI	−8.0484 (10.8010)		−3.1944 (13.2984)	−2.9678 (36.5796)	−1.3866 (16.2100)	−11.3879 (17.1586)
INMETRO	−4.8539 (9.8806)	3.1944 (13.2984)		0.2266 (36.3200)	1.8079 (15.6098)	−8.1935 (16.5904)
UTE	−5.0805 (35.4822)	2.9678 (36.5796)	−0.2266 (36.3200)		1.5812 (37.4879)	−8.4201 (37.9120)
NRC	−6.6618 (13.5466)	1.3866 (16.2100)	−1.8079 (15.6098)	−1.5812 (37.4879)		−10.0013 (18.9898)
CENAM	3.3395 (14.6630)	11.3879 (17.1586)	8.1935 (16.5904)	8.4201 (37.9120)	10.0013 (18.9898)	

lth artefact at different time periods that are all the same for the ith laboratory if $I_i(l) = 1$. From tables 1 and 2, among the five laboratories (NIST, INTI, INMETRO, NRC, CENAM), which made multiple measurments, for a fixed travelling standard only NIST has the same type B standard uncertainty for the five periods. Based on that we assume that $I_1(l) = 1$ for NIST. Otherwise, $I_1(l) = 0$ would lead to different $e_{1j,B}(l)$ and thus possibly different type B uncertainties for the five periods. For the other four NMIs which made measurements in two periods, $I_2(l) = I_3(l) = I_5(l) = I_6(l) = 0$ for $l = 1, 2$. Otherwise, it would contradict the fact that type B uncertainties for a fixed travelling standard and the same NMI are different for the two periods. As for UTE, the fourth laboratory in the tables, $k_4 = 1$ leading to $I_4(l) = 1$.

The slopes of the two regression lines corresponding to the two travelling standards are $\hat{\beta}(1) = 3.6768 \times 10^{-6}$/year and $\hat{\beta}(2) = 4.5873 \times 10^{-6}$/year from (11). For a fixed travelling standard, the intercept of the regression line corresponding to a laboratory is calculated from (10). Figures 1 and 2 show the five regression lines corresponding to the five laboratories with two or more measurements, for S/N 9104 and S/N 9105, respectively. Similar to [3], the CRV as a weighted mean of the measurements calculated from (28) at $\vec{t}^* = (2006.772, 2006.806)$ year with weights given by (26), (27), and (30), is 9.5710×10^{-6}, with a standard uncertainty of 1.6826×10^{-6} from (29).

The degrees of equivalence of the national measurement standards with respect to the CRV and the degrees of equivalence of pairs of national measurment standards and their expanded uncertainties were calculated from (32) to (35) and are listed in tables 3 and 4, respectively.

6. Conclusions

In this paper, we extend the statistical analyses for key comparisons with linear trends in [2, 3] to the general case of possible multiple measurements in multiple time periods by each laboratory. The calculation of the CRV is consistent with the case in [2, 3] and the case in which there is no trend. The corresponding uncertainties for the CRV and the degree of equivalence are also provided.

In this paper, we assume that the measurements of different artefacts by the same laboratory are statistically independent. The assumption is based on (1) the errors quantified by type A uncertainty are statistically independent, (2) the errors quantified by type B uncertainty will have some correlation, and (3) since not all artefacts are created equal and even when metrologists make every effort to measure artefacts in as 'correlated' a way as possible, there is still a random component. Certainly, the scenario that measurements of different artefacts by the same laboratory are correlated can be considered in future research.

Acknowledgments

The authors thank their colleagues W F Guthrie, D G Jarrett and R E Elmquist and two anonymous referees for their helpful comments.

References

[1] International Committee of Weights and Measures 1999 Guideline for CIPM key comparisons (Appendix F to Mutual recognition of national measurement standards and of calibration and measurement certificates issued by national metrology institutes) *Technical Report*

[2] Zhang N F, Liu H-K, Sedransk N and Strawderman W E 2004 Statistical analysis of key comparisons with linear trends *Metrologia* **41** 231–7

[3] Zhang N F, Strawderman W E, Liu H-K and Sedransk N 2006 Statistical analysis for multiple artifact problem in key comparisons with linear trends *Metrologia* **43** 21–6

[4] Elster C, Wöger W and Cox M 2005 Analysis of key comparison data: unstable travelling standards *Meas. Tech.* **48** 883–93

[5] Stepanov A V 2007 General problems of metrology and measurement technique – Comparing algorithms for evaluating key comparisons with linear drift in the reference standard, *Meas. Tech.* **50** 1019–27

[6] Dziuba R F and Jarrett D G 2002 Final report on key comparison CCEM-K2 of resistance standards at 10 MΩ and 1 GΩ *Metrologia* **39** (Tech. Suppl.) 01001

[7] Elmquist R F, Jarrett D G and Zhang N F 2009 RMO comparison final report: 2006–2007 Resistance standards comparison between SIM laboratories SIM.EM-K1, 1 Ω; SIM.EM-K2, 1 GΩ; SIM.EM-S6, 1 MΩ *Metrologia* **46** (Tech. Suppl) 01001

[8] Jarrett D G, Elmquist R F and Zhang N F 2009 SIM comparison of DC resistance standards at 1 Ω, 1 MΩ, and 1 GΩ *IEEE Trans. Instrum. Meas.* **58** 1188–95

[9] Rao C R 2001 *Linear Statistical Inference and its Applications* 2nd edn (New York: Wiley)

[10] NIST International Comparison Database (ICDB) Comparison Information for CCPR-K2.a Draft A

[11] Zhang N F 2006 The uncertainty associated with the weighted mean of measurement data *Metrologia* **43** 195–204