Предлагаем вниманию читателей окончание подборки статей по материалам шестого международного семинара «Математическая, статистическая и компьютерная поддержка качества измерений», проходившего в июне 2012 г. в С.-Петербурге во ВНИИМ. Начало подборки опубликовано в «Измерительной технике» в № 4, 2013 г.

53.081.1

Об одном подходе к оцениванию результатов ключевых сличений эталонов при несогласованных данных

А. Г. ЧУНОВКИНА, Н. А. БУРМИСТРОВА, Н. Д. ЗВЯГИН

Всероссийский научно-исследовательский институт метрологии им. Д. И. Менделеева, С.-Петербург, Россия, e-mail: A.G.Chunovkina@vniim.ru

Рассмотрены методы оценивания данных ключевых сличений и кратко охарактеризованы существующие методы оценивания несогласованных данных. Предложен практический подход к оцениванию таких данных.

Ключевые слова: *степень эквивалентности эталонов, опорное значение ключевых сличений, калибровочные и измерительные возможности, неопределенность измерения.*

The estimation methods of key comparisons data are considered and the existing methods of inconsistent data estimation are briefly characterized. The practical approach to inconsistent data estimation is proposed.

Key words*: degree of standards equivalence, key comparisons reference value, calibration and measurement capabilities, measurement uncertainty.*

На 21-й Генеральной конференции по мерам и весам 14 октября 1999 г. было положено начало подписанию директорами национальных метрологических институтов (НМИ) Договоренности о взаимном признании измерительных эталонов, сертификатов калибровок и измерений (MRA), выдаваемых национальными метрологическими институтами [1]. Объективной основой взаимного признания результатов измерений, проводимых лабораториями НМИ, являются ключевые сличения национальных эталонов и системы менеджмента качества, реализуемые этими институтами.

При помощи ключевых сличений национальных эталонов устанавливают степень их эквивалентности, т. е. степень эквивалентности результатов измерений, получаемых этими НМИ [2]. В соответствии с MRA она выражается двумя значениями: отклонением результата измерения в данной лаборатории от опорного значения ключевого сличения и расширенной неопределенностью этого отклонения. Под опорным значением в большинстве случаев понимается оценка значения измеряемой величины, полученная по результатам всех участников.

В [2] рассмотрена задача оценивания данных ключевых сличений стабильного транспортируемого эталона сравнения при соблюдении ряда дополнительных условий, касающихся независимости измерений в разных лабораториях и нормального закона распределения. Для описанной ситуации модель измерения имеет вид

$$X_i = X, \qquad (1)$$

где X_i — измеряемая величина в i-й лаборатории; участники сличений представляют измеренное значение и соответствующую стандартную неопределенность как x_i, u_i.

Для обработки данных ключевых сличений применяют метод наименьших квадратов [2]. Опорное значение ключевых сличений (KCRV) вычисляют как средневзвешенное x_w результатов измерений в институтах, используя в качестве весов обратные квадраты соответствующих значений стандартной неопределенности:

$$x_w = \frac{x_1/u^2(x_1)+\ldots+x_n/u^2(x_n)}{1/u^2(x_1)+\ldots+1/u^2(x_n)}.$$

При этом стандартная неопределенность будет иметь вид

$$u^2(x_w) = (1/u^2(x_1) + \ldots + 1/u^2(x_n))^{-1}.$$

Средневзвешенное значение может быть взято в качестве KCRV только в том случае, если данные, предоставленные лабораториями, согласуются с принятой моделью (1), которая адекватно описывает данные сличений, что можно проверить по критерию χ^2. Для этого вычисляют наблюдаемое значение статистики

$$\chi^2 = (x_1-y)^2/u^2(x_1) + \ldots + (x_n-y)^2/u^2(x_n).$$

Считается, что проверка согласованности данных не проходит, т. е. данные не подтверждаются, если $P\{\chi^2(n-1) > \chi^2\} < 0{,}05$. В этом случае данные сличений считаются несогласованными.

По мнению авторов, говорить о степени эквивалентности эталонов (эквивалентности результатов измерений) можно только применительно к совокупности согласованных результатов измерений. Однако в MRA нет однозначной трактовки понятия эквивалентности эталонов, устанавливается сразу количественная мера — степень эквивалентности. Нет термина «эквивалентность эталона» и в последней редакции международного словаря по метрологии (VIM 3) [3]. Следовательно, необходимы разъяснения самого понятия «эквивалентность эталонов», его уточнение по отношению к таким понятиям как «точность измерений» и «совместимость результатов измерений».

Степень эквивалентности для согласованных результатов измерений (в случае положительного результата проверки по критерию χ^2) представляет собой пару значений: отклонение результата измерения от опорного значения $d_i = x_i - x_{\text{ref}}$ и неопределенность этого отклонения $u^2(d_i) = u^2(x_i) - u^2(x_{\text{ref}})$.

Результаты сличений и отчеты о них содержатся в Приложении B базы данных BIPM, которая доступна на официальном сайте.

Практическим выходом ключевых сличений является подтверждение их участниками калибровочных и измерительных возможностей (CMC), которое можно представить измеряемыми величинами, диапазонами и расширенной неопределенностью измерений (обычно при уровне доверия 95 %). Их необходимо перечислять по каждому НМИ в Приложении C базы данных BIPM и подтверждать на основе результатов, полученных в ходе ключевых сличений. Это означает, что заявленные CMC неопределенности измерений не могут быть меньше указанных в отчетах по сличениям и, кроме того, результаты сличений должны подтверждать эти неопределенности. Для подтверждения заявленных неопределенностей обычно используют критерий

$$E_n = |x_i - x_{\text{ref}}| \,/\, \left(2u\left(x_i - x_{\text{ref}}\right)\right).$$

Заявляемые неопределенности измерений $u(x_i)$ считаются подтвержденными, если значения $E_n \le 1$. Если $E_n > 1$, то этому может быть две причины:

результат, полученный в лаборатории, существенно отклонился от опорного значения по сравнению с результатами остальных участников;

лаборатория заявила слишком малую неопределенность по сравнению с заявленными неопределенностями остальных участников.

В соответствии с [4] неопределенности, представленные участниками сличений, не могут изменяться в процессе обработки данных. Однако в MRA [1] допускается корректировка CMC по результатам сличений. Такая процедура предполагает, что соответствующая лаборатория проводит анализ бюджета неопределенности, ее возможных неучтенных источников или причин систематического отклонения результата измерений от опорного значения ключевого сличения. Обычно итогом такого анализа является расширение неопределенностей до значений, обеспечивающих соблюдение условия $E_n < 1$.

Рассмотрим далее обработку несогласованных данных сличений. Отметим, что не существует одного признанного метода обработки подобных данных, так как не может быть общего строгого математического обоснования такого метода. Большое число существующих методов [5—14] подразумевает некоторые дополнительные допущения, которые часто носят сугубо математический характер и трудно поддаются практической реализации. Все эти методы следует отнести к эвристическим, которые условно можно разделить на три группы:

исключение «выпадающих» результатов участников и сведение обработки к традиционному подходу. Среди этих алгоритмов следует выделить алгоритм формирования наибольшей согласованной подгруппы результатов измерений [11], когда для проверки согласованности результатов измерений используется критерий χ^2;

изменение модели измерения — введение в нее дополнительных параметров, описывающих неучтенные систематические смещения данных лабораторий (fixed effect model) или их дополнительный случайный разброс между лабораториями (random effect model) [8—10, 12—14]. Получение оценок этих параметров требует дополнительных предположений;

расширение (увеличение) неопределенностей исходных данных и, возможно, «исправление» самих данных так, чтобы они стали согласованными (см., например [5, 6]).

Предлагаемый ниже алгоритм может быть отнесен к последней группе и подразумевает расширение неопределенностей измерений для ряда участников. При этом эквивалентность эталонов понимается, прежде всего, как согласованность результатов измерений, выполняемых лабораториями — участницами сличений, и их совместимость с опорным значением ключевых сличений. В этом контексте расширение неопределенностей трактуется как достижение определенного уровня согласованности результатов внутри некоторой группы лабораторий, для которой устанавливается эквивалентность результатов измерений. Расширение неопределенностей касается только тех результатов измерений, заявленные неопределенности которых представляются заниженными по сравнению с отклонениями этих результатов от опорного значения, т. е. эти результаты несовместимы с опорным значением ключевых сличений. В определенном смысле предлагаемая процедура расширения неопределенности измерений соответствует ее коррекции, выполняемой по результатам сличений теми участниками, которые не подтвердили в ходе сличений заявляемые неопределенности.

Алгоритм опирается на базовую модель (1): каждая лаборатория измеряет одну и ту же величину X и предоставляет результат измерения X_i с соответствующей стандартной неопределенностью u_i, $i = 1, \ldots, n$. Априори предполагается, что данные не согласованы, т. е. выборочное значение статистики χ^2 превышает критическое для уровня значимости 5 %. Как правило, для ряда участников сличений оказывается $E_n > 1$, что свидетельствует о несовместимости этих результатов измерений с опорным значением ключевых сличений.

Оба последних обстоятельства показали, что при формировании бюджета неопределенности ряд участников сличений допустили ошибки. Поэтому ставим задачу предложить такие оценки неопределенности, которые позволили бы сделать данные согласованными или, что равнозначно, сде-

лать результаты измерений эквивалентными. Расширение неопределенности предполагаем только для тех участников, для которых значения критерия E_n оказываются больше единицы.

Особо отметим, что достижение эквивалентности результатов измерений не означает, что все результаты будут иметь одну и ту же степень эквивалентности, которая характеризуется, прежде всего, неопределенностью измерений, связанной с отклонением результата от опорного значения. Уровень точности определяется именно этой неопределенностью: чем она меньше, тем выше точность измерений в данной лаборатории. Отклонение результата измерения от опорного значения важно только в качестве объективного подтверждения этого уровня точности.

Предлагаем следующую процедуру расширения неопределенности измерений для части участников сличений, результаты которых оказались несовместимыми с опорным значением. Сначала выделим референтную группу результатов измерений, которая представляет собой однородную совокупность без явных выбросов. Для формирования этой группы будем использовать критерий z_n. Таким образом, в референтную группу будут входить результаты измерений, для которых $z_n < 1$:

$$z_n = \frac{|x_i - \bar{x}|}{2S};\ \bar{x} = \frac{1}{m}\sum_{i=1}^{m} x_i;\ S^2 = \frac{1}{m-1}\sum_{i=1}^{m}(x_i - \bar{x})^2,$$

где m — число результатов в референтной группе.

Выборочное среднее квадратическое отклонение (СКО), вычисленное по результатам измерений, вошедшим в референтную группу, в качестве расширенного значения неопределенности предлагаем для тех участников, результаты которых несовместимы с опорным значением $u = S$. После этого процедуру оценивания данных повторяем: проверяем согласованность данных по критерию χ^2 и, если он проходит, вычисляем опорное значение ключевых сличений, а если не проходит, то исключаем результаты измерений, несовместимые с опорным значением (критерий E_n). Если критерий E_n (с увеличенными неопределенностями) не прошел, это означает, что соответствующий результат значительно смещен относительно опорного значения и для него степень эквивалентности не определяется. Для остальных результатов измерений устанавливаем степень эквивалентности с учетом измененных неопределенностей.

В качестве примера рассмотрим оценивание результатов сличений CCM.P-K12. В настоящее время в базе данных BIPM опубликован отчет об этих сличениях. Поскольку результаты сличений оказались несогласованными, то при их оценивании были применены различные подходы. Один из них, в частности, представлен в [15]. Здесь приведем результаты использования предлагаемого алгоритма к данным сличениям.

В сличениях CCM.P-K12 измеряемой величиной является поток газа в вакууме. Поскольку изначально предполагался систематический дрейф измеряемой величины в процессе транспортировки эталона сравнения, была выбрана радиальная схема сличений. При этой схеме пилот-лаборатория неоднократно выполняла измерения в ходе сличений, что позволило оценить параметры дрейфа с высокой точностью и внести поправки в результаты измерений. В табл. 1 представлены скорректированные данные сличений после исключения дрейфа эталона сравнения и значения критерия E_n. В качестве опорного было выбрано средневзвешенное значение с весами, обратно пропорциональными квадратам заявленных стандартных неопределенностей.

Таблица 1

Результаты измерений участников сличений

НМИ	$q_v \cdot 10^{11}$, моль/с	$u \cdot 10^{13}$, моль/с	E_n
1	4,3355	1,7	0,99
2	4,3135	7,5	0,35
3	4,4283	5,9	0,53
4	4,7067	9,2	1,85
5	4,3334	1,8	0,98
6	4,2250	5,2	1,38
7	4,2779	3,4	1,33
8	4,3050	4,3	0,73
9	4,4280	1,2	3,07
10	4,3347	1,6	1,07
11	4,4558	5,3	0,84

Графически исходные данные представлены на рис. 1.

При проверке согласованности результатов сличений CCM.P-K12 использовали критерий χ^2, который показал несогласованность данных. Выборочное его значение составило 70,26, при этом предельно допустимое — 18,31.

В референтную группу вошли результаты всех участников, кроме заявленного НМИ 4. Значение неопределенности, рассчитанное как выборочное СКО результатов референтной группы, составило $7{,}3 \cdot 10^{-13}$. Сопоставление со значениями неопределенностей, заявленными участниками сличений (см. табл. 1), показывает, что только три лаборатории представили существенно меньшую неопределенность.

После расширения неопределенностей для участников референтной группы (НМИ 6, 7, 9 и 10), для которых $E_n > 1$, до установленного уровня $7{,}3 \cdot 10^{-13}$ результаты этой группы стали согласованными по критерию χ^2 ($\chi^2 = 12{,}71$, при этом предельно допустимое значение 16,92). Пересчитанное опорное значение $x_{ref} = 4{,}362 \cdot 10^{-11}$ с соответствующей неопределенностью $u_{ref} = 1{,}06 \cdot 10^{-13}$. Первоначально опорное значение было существенно смещено в сторону результата НМИ 9, который имел очень малую неопределенность (при-

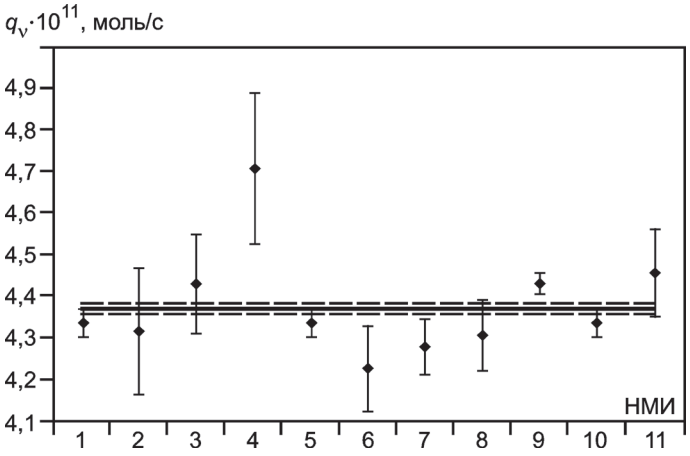

Рис. 1. Исходные данные сличений CCM.P-K12

Таблица 2

НМИ	1	2	3	4	5	6	7	8	9	10	11
E_n	0,1	0,13	0,82	2,05	0,03	0,74	0,84	0,33	0,66	0,08	0,85

чем не подтвержденную результатами измерений). Расширение неопределенности измерений НМИ 9 привело к изменению опорного значения. Теперь неопределенности измерений НМИ 7 и 10 не нуждаются в изменениях, но должна быть расширена неопределенность измерения для НМИ 11.

После этих изменений вновь проверили согласованность по критерию χ^2, значение которого составило 12,53 при предельно допустимом значении 16,92, уточнили опорное значение ($x_{ref} = 4{,}341 \cdot 10^{-11}$, $u_{ref} = 8{,}78 \cdot 10^{-14}$) и вычислили E_n (табл. 2).

Таким образом, все результаты (за исключением результата НМИ 4, который значительно отличается от остальных) согласованы и НМИ подтвердили заявленные неопределенности с учетом предложенного расширения для трех НМИ. Теперь участники сличений могут заявлять указанные неопределенности измерений в калибровочных и измерительных возможностях.

Результаты применения данного алгоритма к обработке сличений CCM.P-K12 представлены на рис. 2 для первой течи (именно для нее результаты измерений первоначально были несогласованными). Темными квадратами отмечены те лаборатории, неопределенности измерений которых были расширены (это НМИ 6, 9 и 11). Видно, что все результаты, за исключением одного, совместимы с опорным значением. Более того, расширенные неопределенности не выделяются как неправдоподобно большие по сравнению с неопределенностями остальных участников.

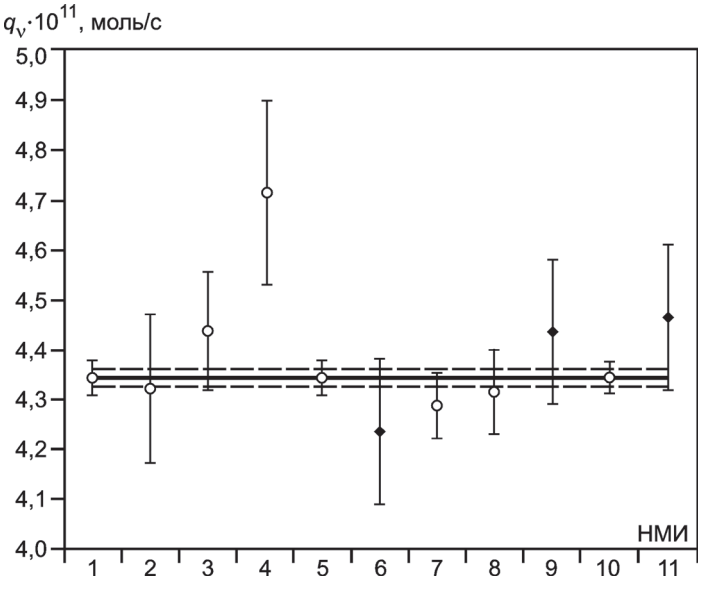

Рис. 2. Результаты обработки данных сличений CCM.P-K12 в соответствии с предлагаемым алгоритмом

В табл. 3 представлены степени эквивалентности результатов сличения участников кроме результата НМИ 4, для которого степень эквивалентности не может быть установлена, поскольку он не согласован с опорным значением.

Таблица 3

Степени эквивалентности результатов измерений для участников сличения CCM.P-K12

НМИ	$D_{1j} \cdot 10^{13}$, моль/с	$U_{D1j} \cdot 10^{13}$, моль/с	НМИ	$D_{1j} \cdot 10^{13}$, моль/с	$U_{D1j} \cdot 10^{13}$, моль/с
1	0,2822	2,86	7	–5,4569	6,48
2	–1,907	14,9	8	–2,7886	8,42
3	9,605	11,8	9	9,500	14,5
5	0,0870	3,15	10	0,2140	2,82
6	–10,721	14,5	11	12,353	14,5

Таким образом, предлагаемый подход к оцениванию несогласованных данных ключевых сличений ориентирован на трактовку понятия эквивалентности эталонов (результатов измерений) как совместимости результатов измерений с опорным значением ключевых сличений. Опорное значение при этом понимается как оценка значения измеряемой величины, полученная по согласованной группе результатов измерений. Подход предполагает увеличение неопределенностей измерений тех результатов, которые оказались несовместимыми с опорным значением. Расширение неопределенности осуществляется до некоторого уровня, который устанавливается по разбросу данных референтной подгруппы, образующих однородную совокупность.

Литература

1. **BIPM 1999.** Mutual Recognition of National Measurement Standards and of Calibration and Measurement Certificates issued by National Metrology Institutes (MRA).

2. **Cox M. G.** The evaluation of key comparison data // Metrologia. 2002. V. 39. P. 589—595.

3. **ISO/IEC Guide 99:2007.** International vocabulary of metrology — Basic and general concepts and associated terms (VIM).

4. **CIPM MRA-D-05.** Measurement comparisons in the CIPM MRA [Офиц. сайт]. http: // www.bipm.org (дата обращения 28.11.2012).

5. **Weise K., Woger W.** Removing model and data nonconformity in measurement evaluation // Meas. Sci. Technol. 2000. N 11. P. 1649—1658.

6. **Willink R.** Statistical determination of a comparison reference value using hidden errors // Metrologia. 2002. V. 39. P. 343—354.

7. **Нильсен Л.** Идентификация и обращение с отличными от основной группы результатами измерений при ключевых сличениях // Измерительная техника. 2003. № 5. С. 63—68; **Nielsen L**. Identification and Handling of Discrepant Measurements in Key Comparisons // Measurement Techniques. 2003. V. 46. N 5. P. 513—522.

8. **White D. R.** On the analysis of measurement comparisons // Metrologia. 2004. V. 41. P. 122—131.

9. **Kacker R**. Combining information from interlabaratory evaluations using a random effects model // Ibid. P. 132—136.

10. **Sutton C. M.** Analysis and linking of international measurement comparisons // Ibid. P. 272—277.

11. **Cox M. G.** The evaluation of key comparison data: determining the largest consistent subset // Metrologia. 2007. V. 44. P. 187—200.

12. **Lira I.** Combining inconsistent data from interlaboratory comparisons// Ibid. P. 415—421.

13. **Elster C., Toman B.** Analysis of key comparisons: estimating laboratories' biases by a fixed effects model using Bayesian model averaging // Metrologia. 2010. V. 47. P. 113—119.

14. **Toman B., Possolo A.** Laboratory effects models for interlaboratory comparisons // Accreditation and Quality Assurance. 2009. V. 14. P. 553—563.

15. **Боднар О. и др.** Применение техники байесовского усреднения до модели с фиксированными эффектами и линейным трендом для анализа ключевого сличения CCM.P-K12 // В этом номере журнала. С. 7—11.

Дата принятия 28.01.2013 г.

389.17

Применение техники байесовского усреднения для модели с фиксированными эффектами и линейным трендом при анализе ключевого сличения CCM.P-K12

О. БОДНАР, А. ЛИНК, К. КЛАУЭНБЭРГ, К. ЖУСТЭН, К. ЭЛЬСТЭР

Государственный физико-технический институт, Берлин, Германия,
e-mail: olga.bodnar@ptb.ru

Выполнено обобщение процедуры оценивания данных на случай линейного дрейфа характеристик транспортируемого эталона. Процедура основана на методе байесовского усреднения моделей с фиксированными эффектами в предположении, что некоторое число лабораторий измеряет без смещения. Метод применен для оценивания данных ключевого сличения CCM.P-K12.

Ключевые слова: *модель с фиксированными эффектами, байесовское усреднение, взвешенный метод наименьших квадратов, смещение.*

The procedure of data estimation is generalized for the case of linear drift characteristics of transported standard. The procedure is based on Bayesian averaging of models with fixed effects in assumption that a certain number of laboratories measure without bias. The method has been applied for the key comparison CCM.P-K12 data estimation.

Key words: *fixed effects model, Bayesian averaging, weighted least squares method, bias.*

Для анализа ключевых сличений в [1] был предложен метод байесовского усреднения моделей с фиксированными эффектами. В статье данный метод обобщен на случай линейного дрейфа характеристик транспортируемого эталона и использован для анализа данных текущего ключевого сличения CCM.P-K12. Оно проводится в области измерений давлений и направлено на проверку калибровочных и измерительных возможностей использования потока газа в вакууме, т. е. очень низких скоростей потока [2, 3]. В этом ключевом сличении приняли участие 11 национальных метрологических институтов (НМИ). Измеряемой величиной является поток гелия, истекающего в вакуум из объема транспортируемого эталона. Пилотная лаборатория проводила дополнительные контрольные измерения до, в течение и после сличения для определения зависящего от времени дрейфа измеряемой величины. На рис. 1 показаны результаты измерений всех участников в зависимости от времени, а также контрольных измерений пилотной лаборатории (PTB). Наличие линейного дрейфа можно также предположить из результатов измерений всех лабораторий, но независимо от линейного дрейфа данные оказались несовместимыми. Чтобы иметь равный вес среди всех участников для последующего анализа, рассматривали только одно (первое) измерение пилотной лаборатории.

Ниже данные сличения проанализированы в соответствии с предложенным методом, основанным на модели с фиксированными эффектами результатов измерений лабораторий. Эта модель содержит следующие неизвестные параметры: опорное значение результата ключевого сличения и смещение результата измерения для каждой лаборатории. Целью является оценка этих смещений, которые отражают упущенные из рассмотрения возможные эффекты, не вошедшие в бюджет неопределенности. Предлагаемый метод основан на предположении, что можно указать количество лабораторий (нижнюю границу), для которых смещение равно нулю. Некоторые рекомендации по определению количества таких лабораторий можно получить из данных сличений. Далее рассмотрены различные модели, где каждая модель включает разные подмножества лабораторий, которые, как предполагается, имеют нулевое смещение. На-