

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 ССМ.Р-К12 // 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 ССМ.Р-К12

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 ССМ.Р-К12.

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 ССМ.Р-К12 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 ССМ.Р-К12. 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 (РТВ). Наличие линейного дрейфа можно также предположить из результатов измерений всех лабораторий, но неза-

висимо 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-

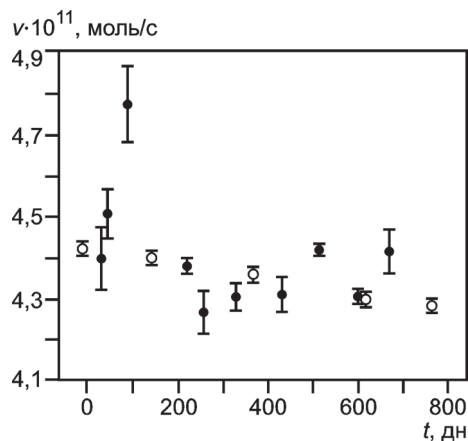


Рис. 1. Данные сличения ССМ.Р-К12:

черные точки — результаты измерений скорости потока газа десятию НМИ и их стандартные неопределенности; светлые точки — результаты контрольных измерений пилотной лаборатории (РТВ)

конец, результаты для всех этих моделей объединяются с использованием байесовского метода усреднения [4], что дает опорное значение и оценки смещения для каждой лаборатории. Полученные оценки смещения предлагается рассматривать как степени эквивалентности (DOE) для ключевых сличений [5].

Одно из преимуществ метода состоит в том, что даже при несовместимых (несогласованных) данных результаты (без исключения и существенно отклоняющихся) всех лабораторий учитываются при вычислении опорного значения и степеней эквивалентности, а возможные погрешности измерений изначально включены в модель. Кроме этого, данный метод преимущественно подходит для случая, когда участвующие лаборатории предоставляют распределения вероятностей, суммирующие состояние знаний о значении измеряемой величины [1], и соответствует методу оценки неопределенности [6].

Метод, предложенный в [1], предполагает, что транспортируемый эталон стабилен, что не всегда верно. Цель данной работы — обобщить метод на случай линейного дрейфа, который ожидается исходя из физических закономерностей, и применить его для анализа данных ключевого сличения ССМ.Р-К12. Как и в [1], ограничимся случаем, когда можно предположить гауссово распределение для измеряемых величин со средними значениями и стандартными отклонениями, равными измеренным значениям и соответствующим стандартным неопределенностям, представленными лабораториями-участниками. В статье проводится детальное обсуждение и оценка результатов ключевых сличений ССМ.Р-К12, включая интерпретацию полученных оценок смещений. Одним из результатов обобщенного метода является оценка дрейфа. Поскольку в ключевом сличении ССМ.Р-К12 пилотная лаборатория представила результаты повторных контрольных измерений, их можно использовать для независимой оценки дрейфа и проверки согласованности двух его оценок.

Теоретические положения

Байесовская модель усреднения для модели с фиксированными эффектами. Обозначим оценки и соответствующие стандартные неопределенности как вектор-строку $x = (x_1, \dots, x_n)^T$ и $u = (u(x_1), \dots, u(x_n))^T$, где n — количество лабо-

раторий, участвующих в ключевом сличении. Результаты измерений $x_1, \dots, x_n$ моделируются как реализации случайных величин

$$X_i = \mu + \beta t_i + \alpha_i + E_i, \quad i = 1, \dots, n, \quad (1)$$

где $\mu + \beta t_i$ — значение измеряемой величины в момент времени t_i ; $t_1, \dots, t_n$ — время (известное) проведения измерений; α_i — смещение результатов i -й лаборатории; $E_i \sim N(0, u^2(x_i))$ — погрешности, нормально распределенные с нулевым средним и дисперсией $u^2(x_i)$.

Согласно [1] для обоснования вида распределения погрешностей E_i предполагаем, что гауссово распределение (для независимых величин) со средним x_i и дисперсией $u^2(x_i)$ характеризует состояние знаний о значении измеряемой величины для каждой лаборатории. Модель (1) содержит $n+2$ неизвестных параметров: μ , β и $\alpha_1, \dots, \alpha_n$. Поскольку доступно только n наблюдений $x_1, \dots, x_n$, модель (1) недоопределена. Чтобы перейти к четко определенной задаче, применяется то же самое дополнительное ограничение, как и в [1]:

$$\alpha_{l_i} = 0, \quad l_i \in \{l_1, \dots, l_m\},$$

где m — параметр порядка модели, откуда следует, что для неизвестного множества лабораторий были предположительно указаны достоверные неопределенности, т. е. соответствующие смещения равны нулю. Чтобы модель была определена, требуется $m \geq 2$. Далее предполагается, что параметр m зафиксирован. Это означает, что результаты измерений m лабораторий не имеют смещения, но не указывается, какие именно.

Анализ проводится по аналогии с [1]. Рассматриваются все возможные модели M_I , где $I = \{l_1, \dots, l_m\}$ — множество

лабораторий, для которых соответствующие смещения α_{l_i} , как предполагается, равны нулю. Например, для $m = 3$ и $I = \{1, 2, 3\}$ смещения $\alpha_1, \alpha_2, \alpha_3$ будут равны нулю и, следовательно, исключены из (1). Для этой модели уравнения будут выглядеть таким образом:

$$X_1 = \mu + \beta t_1 + E_1; \quad X_2 = \mu + \beta t_2 + E_2; \quad X_3 = \mu + \beta t_3 + E_3;$$

$$X_4 = \mu + \beta t_4 + \alpha_4 + E_4; \dots$$

Параметры модели μ , β , $\alpha_4, \alpha_5, \dots$ — неизвестны, и априори о них нет никаких знаний. Поэтому используются неинформативные априорные плотности распределений вероятностей в виде констант, так как эти параметры входят в функцию правдоподобия линейно. При помощи байесовского метода находим соответствующие апостериорные распределения. На основе последовательного анализа всех моделей M_I , каждая из которых содержит ровно m лабораторий, имеющих нулевое смещение результатов измерений,

получаем $\frac{n!}{m!(n-m)!}$ (разных) апостериорных распределений, которые объединяются с помощью метода байесовского усреднения¹ [4]. Обоснование его применения в соче-

¹ В принципе, вычисления ведутся аналогично [1]. Разница в том, что конструкция матрицы в приложении к [1] обобщается для модели с линейным дрейфом и соответствующие аналитические формулы, выведенные в [1], не могут быть больше использованы, поэтому, например, детерминанты матриц находятся численно.

тании с неинформативными априорными плотностями распределений для набора связанных моделей той же размерности приведены в [1]. После интегрирования по мешающим параметрам получаем распределения параметров, которые обозначим как $p(\mu|x, m)$, $p(\beta|x, m)$, $p(\alpha_1|x, m)$, ..., $p(\alpha_n|x, m)$.

Оценку $\hat{\mu}$, где μ — значение измеряемой величины в момент времени $t = 0$, вычисляем как математическое ожидание соответствующего апостериорного распределения:

$$\hat{\mu} = \int_{-\infty}^{\infty} \mu p(\mu|x, m) d\mu,$$

ее стандартная неопределенность

$$u^2(\hat{\mu}) = \int_{-\infty}^{\infty} (\mu - \hat{\mu})^2 p(\mu|x, m) d\mu.$$

Аналогично можно вычислить оценки $\hat{\beta}$ и $\hat{\alpha}_1, \dots, \hat{\alpha}_n$ вместе с соответствующими стандартными неопределенностями для скорости β изменения измеряемой величины и смещений $\alpha_1, \dots, \alpha_n$ для отдельных лабораторий. Доверительные интервалы (интервалы охвата в [6]), также можно вычислить, используя полученные распределения. Например, симметричный относительно оценки 95 %-й доверительный интервал задается как

$$[\hat{\alpha}_1 - U(\hat{\alpha}_1), \hat{\alpha}_1 + U(\hat{\alpha}_1)], \quad (2)$$

где $U(\hat{\alpha}_1)$ выбрано так, что верно соотношение

$$\int_{\hat{\alpha}_1 - U(\hat{\alpha}_1)}^{\hat{\alpha}_1 + U(\hat{\alpha}_1)} p(\alpha_1|x, m) d\alpha_1 = 0,95.$$

Далее будем использовать симметричные доверительные интервалы вида (2).

Другой подход к определению доверительного интервала заключается в нахождении наикратчайшего интервала. Однако это может оказаться неудачным, поскольку область, где плотность апостериорного распределения достигает максимумов, может разбиться на две подобласти, всегда включающие нуль. Причина в том, что распределения смещений являются сочетанием дискретных и непрерывных

распределений, где дискретная часть с ненулевым значением сосредоточена в нуле (см. также соответствующее обсуждение в [1]).

Выбор порядка модели. Порядок m модели невозможно получить из результатов измерений, но его можно установить. К счастью, даже в отсутствие какой-либо дополнительной информации, m не может превышать максимального числа $\bar{m}$ лабораторий, которые образуют когерентное подмножество. Подмножество лабораторий называется когерентным, если результаты любой пары лабораторий в подмножестве существенно не отличаются друг от друга. Другими словами, если любое подмножество результатов измерений k лабораторий не является согласованным в рамках заявленных неопределенностей с учетом линейного дрейфа (но без учета смещений для каждой лаборатории), то должно быть $m < k$. Также при определении m можно использовать и другую информацию, например, когда вероятно разбиение лаборатории на группы в соответствии с использованием различных методов измерения и при систематических расхождениях между этими методами.

В [1] верхняя граница $\bar{m}$ была получена как число лабораторий в наибольшем когерентном подмножестве лабораторий. Однако способ ее определения, который представлен в [1], нельзя применить в данном случае при наличии дрейфа. Поэтому обобщим метод определения $\bar{m}$ следующим образом. Сначала рассмотрим все тройки лабораторий и проверим, показывают ли они результаты, которые согласуются с моделью линейного тренда. Определим тестовую статистику для любых трех лабораторий (i_1, i_2, i_3) :

$$T_{i_1 i_2 i_3} = \left(\begin{pmatrix} x_{i_1} \\ x_{i_2} \\ x_{i_3} \end{pmatrix} - \begin{pmatrix} 1 & t_{i_1} \\ 1 & t_{i_2} \\ 1 & t_{i_3} \end{pmatrix} \begin{pmatrix} \hat{\mu} \\ \hat{\beta} \end{pmatrix} \right)^T \begin{pmatrix} u^2(x_{i_1}) & 0 & 0 \\ 0 & u^2(x_{i_2}) & 0 \\ 0 & 0 & u^2(x_{i_3}) \end{pmatrix}^{-1} \times \\ \times \left(\begin{pmatrix} x_{i_1} \\ x_{i_2} \\ x_{i_3} \end{pmatrix} - \begin{pmatrix} 1 & t_{i_1} \\ 1 & t_{i_2} \\ 1 & t_{i_3} \end{pmatrix} \begin{pmatrix} \hat{\mu} \\ \hat{\beta} \end{pmatrix} \right), \quad (3)$$

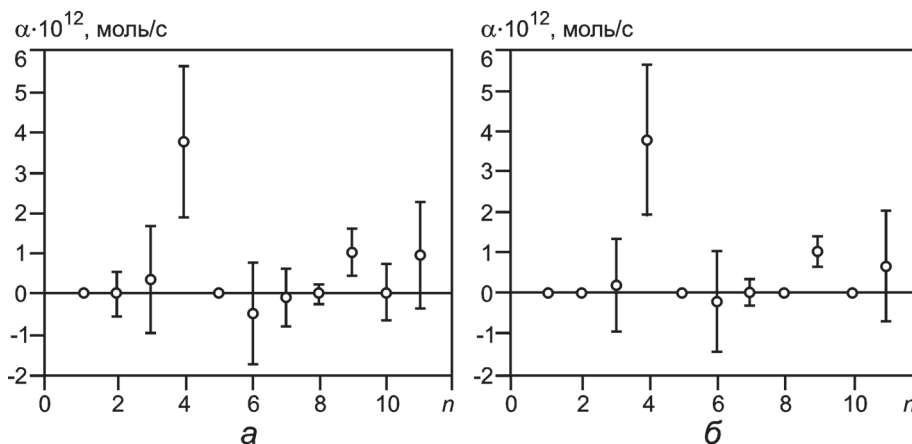


Рис. 2. Оценки смещения вместе с их 95 %-ми симметричными доверительными интервалами по данным рис. 1 для порядков модели $m = 7$ (а) и $m = 8$ (б)

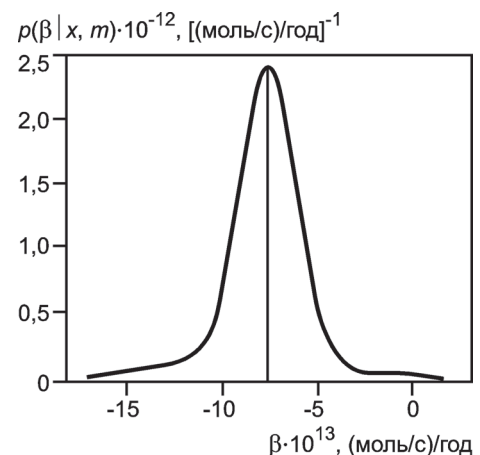


Рис. 3. Распределение скорости β изменения измеряемой величины, полученное по данным рис. 1 для пилотной лаборатории; $m = 7$

которая в случае согласованности будет иметь χ^2_1 -распределение, т. е. $T_{i_1 i_2 i_3} \sim \chi^2_1$. В (3) $\hat{\mu}, \hat{\beta}$ обозначают оценки μ, β , полученные взвешенным методом наименьших квадратов для трех конкретных лабораторий (i_1, i_2, i_3). Поскольку существует $L = \frac{n!}{3!(n-3)!}$ троек, контролируем совместную ошибку 1-го рода, выбрав критическое (верхнее) значение для каждой $T_{i_1 i_2 i_3}$, которое равно $\chi^2_{1; 1-\alpha/L}$ (квантиль χ^2 -распределения с одной степенью свободы при уровне значимости $1 - \alpha/L$), где $\alpha = 0,05$. Далее рассмотрим все когерентные подмножества этих лабораторий. Назовем подмножество когерентным, если все тройки лабораторий в нем соответствуют модели линейного тренда, как описано выше. Верхняя граница $\bar{m}$ для m тогда определяется как число лабораторий в наибольшем когерентном подмножестве.

Можно также проанализировать результаты в зависимости от m , где $2 \leq m \leq \bar{m}$, после чего выбрать такое m , для которого результаты являются наименее чувствительными к его изменению. Не рекомендуется выбирать $m = \bar{m}$, поскольку в этом случае при определении весов в байесовском методе усреднения может доминировать единственное когерентное подмножество. Если желательно автоматически фиксировать значение m , то можно установить некоторое правило. Например, выбрать наименьший порядок модели $m \leq \bar{m}$, при котором для наибольшего числа лабораторий оценки смещений значимы. Для ключевого сличения ССМ.Р-К12 это верно, когда $m = 7$.

Результаты исследования. В соответствии с процедурой выбора m для ключевого сличения ССМ.Р-К12 была установлена верхняя граница $\bar{m} = 9$. При выборе $\bar{m}$ никакая дополнительная информация не используется. Параметр порядка модели был найден автоматически, как описано выше, и в этом случае $m = 7$. На рис. 2 показаны оценки смещения для каждой лаборатории и 95 %-е доверительные интервалы для моделей двух разных порядков $m = 7$ и 8. Проведение анализа также для $m = 8$ показало, что значимые смещения нечувствительны к такому выбору порядка модели. Сравнивая оценки на рис. 2, полученные для $m = 7$

и 8, наблюдаем изменения ширины доверительных интервалов для некоторых α_i . Причина в том, что увеличение m на единицу означает предположение, что результаты измерений еще одной лаборатории не имеют смещения, что приводит к изменениям весов в методе байесовского усреднения.

Всего две оценки смещения нельзя признать значимыми, т. е. их 95 %-е доверительные интервалы не содержат нуль. Если предположить, что, по крайней мере, $m = 7$ лабораторий представили достоверные неопределенности (и предположение модели (1) верно), можно сделать следующий вывод: соответствующие лаборатории упустили из виду некоторые источники неопределенности в бюджетах неопределенности.

Другим результатом является апостериорное распределение $p(\beta|x, m)$ для скорости β изменения измеряемой величины (рис. 3). Также интересно, каково значение β , если оценивать его, используя только результаты контрольных измерений пилотной лаборатории (см. рис. 1).

Взвешенный метод наименьших квадратов приводит к $\hat{\beta} = -7,57 \cdot 10^{-13}$ (моль/с)/год. Оценка для β , полученная с использованием апостериорного распределения $p(\beta|x, m)$, $\hat{\beta} = -7,64 \cdot 10^{-13}$ (моль/с)/год. Оба значения очень близки друг к другу (на рис. 3 неразличимы) и к пику функции плотности.

Обсуждение и выводы. В статье представлен метод анализа данных ключевых сличений для случая, когда предполагается наличие дрейфа характеристик транспортируемого эталона, описываемого линейной функцией. Если предположить, что результаты измерений, по крайней мере, m лабораторий не имеют смещения, то метод байесовского усреднения, применяемый к моделям с систематическими смещениями, позволяет выявить и оценить линейный дрейф, а также смещение от опорного значения для каждой лаборатории.

Результаты при применении предлагаемого метода зависят от выбора параметра порядка модели m . Как уже отмечалось в [1], эта зависимость от m является недостатком метода, а следствием того, что проблема оценки смещения для лабораторий в ходе ключевого сличения оказывается некорректной и может быть решена только путем добавления информации о смещениях для лабораторий или измеряемой величины. Полагаем, что предположение о количестве лабораторий, которые предоставляют достоверные неопределенности (без указания этих лабораторий), является допустимым в метрологии. Уменьшение значений параметра порядка модели m означает более осторожное допущение о числе таких лабораторий. Соответственно результаты анализа данных сличения становятся более консервативными, т. е., как правило, неопределенности, связанные с оценками смещений, увеличиваются, если уменьшается m . Например, на рис. 4, а показаны оценки смещений с 95 %-ми доверительными интервалами для девятой лаборатории-участницы ССМ.Р-К12 в зависимости от m . Ширина 95 %-х доверительных интервалов монотонно возрастает, когда m уменьшается, и для $m < 7$ оценки смещения уже

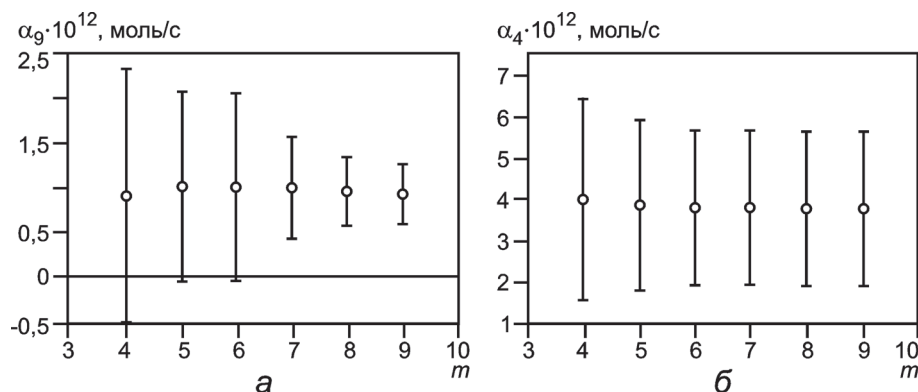


Рис. 4. Оценки смещений с их 95 %-ми симметричными доверительными интервалами для девятой (а) и четвертой (б) лабораторий по данным рис. 1 в зависимости от порядка m модели

не будут значимыми. Это показывает, что возможно данная лаборатория упускает из виду некоторые систематические эффекты при оценивании неопределенности. Однако окончательный вывод может быть сделан только при допущении, что, по крайней мере, $m = 7$ лабораторий предоставляют достоверные неопределенности.

Напротив, значимыми являются все оценки смещений для четвертой лаборатории (см. рис. 4, б) независимо от модели в интервале $m = 4 \dots 9$. Таким образом, это означает, что четвертая лаборатория упускает из виду некоторые эффекты в оценке неопределенности.

В соответствии с [5] цель ключевых сличений заключается в оценке калибровочных и измерительных возможностей СМС, зарегистрированных в базе данных BIPM. Если некоторые оценки смещений являются значимыми, то необходимо провести дальнейшее исследование, и, возможно, соответствующие СМС должны быть пересмотрены. По рекомендациям [5] степени эквивалентности эталонов следует рассчитывать как разности между результатами измерений и опорным значением. Однако оценки смещений, полученные описанным методом, не вычисляются иначе. Тем не менее, смещения для лабораторий (в смысле упущенных из виду эффектов, которые не учитываются в бюджете неопределенности) можно рассматривать как величины, которые должны оценивать степени эквивалентности [5]. Эти величины характеризуют эффекты, которые не были учтены лабораториями. Если соответствующие оценки смещения являются значимыми, то необходимо провести дальнейшие исследования, и, возможно, соответствующие СМС должны быть пересмотрены. Оказалось, что для предлагаемого статистического метода, а также для других методов [9—11], оценки смещения не равны разнице между результатами измерений и опорным значением.

Наконец, следует отметить, что анализ данных ключевых сличений никогда не может доказать правильность измерений, на его основе можно только сделать выводы о согласованности или когерентности результатов лабораторий-участниц. В случае несогласованных данных необходима дополнительная информация или предположения для того, чтобы сделать выводы. Например, даже предположение, что некоторые из лабораторий предоставляют достоверные неопределенности (как это требуется для обсуждаемого метода), может оказаться неправильным, и одна единствен-

ная лаборатория, результаты измерений которой не согласованы с данными других лабораторий, на самом деле предоставляет правильные результаты. Чтобы исключить такую возможность, нужно было бы проанализировать все эксперименты и понять физическую причину наблюдаемой несогласованности данных. Если это невозможно, то приходится вводить дополнительные предположения о данных сличений. Считаем, что предположения, которые приняты в рассматриваемом методе, естественны и приемлемы для метрологов.

Литература

1. **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.
2. **Ehrlich C. D.** A note on flow rate and leak rate units // *J. Vacuum Sci. & Technol.* 1986. V. A4. P. 2384—2385.
3. **Jitschin W., Grosse G., Wandrey D.** Diffusion leak artifacts as a secondary standard for gas flow // *Vacuum*. 1988. V. 38. P. 883—886.
4. **Hoeting J. A. e. a.** Bayesian Model Averaging: A Tutorial // *Statistic. Sci.* 1999. V. 14. P. 382—417.
5. **Mutual Recognition of National Measurement Standards and of Calibration and Measurement Certificates issued by National Metrology Institutes (MRA).** BIPM, 1999.
6. **JCGM 101:2008 BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, OIML.** Evaluation of measurement data. Suppl. 1 to the «Guide to the expression of uncertainty in measurement» Propagation of distributions using a Monte Carlo method.
7. **Kass R. E., Wasserman L.** The selection of prior distributions by formal rules // *J. Am. Stat. Assoc.* 1996. V. 91. P. 1343—1370.
8. **Harville D. A.** Matrix Algebra from a Statistician's Perspective. N.Y.: Springer-Verlag, 1997.
9. **Chunovkina A. G., Elster C., Lira I., Wöger W.** Analysis of key comparison data and laboratory biases // *Metrologia*. 2008. V. 45. P. 211—216.
10. **Sutton C. M.** Analysis and linking of international measurement comparisons // *Metrologia*. 2004. V. 41. P. 272—277.
11. **White D. R.** On the analysis of measurement comparisons // *Ibid.* P. 122—131.

Дата принятия 09.01.2013 г.