# 标准参考物质在分析数据
# 及分析方法检验中的应用

**前言**

Certified Reference Material（标准参考物质，下简称 CRM）在分析化学中的重要性已被人们日趋认识。在欧美国家大约已有130家各种类型CRM的生产者。如果仅就成份CRM而言,如矿砂、岩石及合金等,至少有39家生产单位总共提供大约250种标准矿砂、精矿、岩石、土壤及水泥,以及2400种以上的金属或合金的CRM。

CRM广泛用于分析数据的质量控制上,利用它们可直接校正方法和仪器,检验用于控制分析方法的精密度、准确度及灵敏度,特别是在仲裁分析时,CRM作为定值的依据,对于提高分析水平,保证产品质量起着极大的作用。法国的Oster等人[1]及日本的Mashiko等人[2]研究了CRM在仪器和方法直接校正方面的应用,CCRMP(加拿大CRM计划)[3]根据13年的标准矿砂和相关物质制造经验,提出了在分析方法准确度和精密度的测定过程中应用CRM的实验设计。本文[4]主要介绍这一实验设计,以及对证书值固有的不确定度的统计传递。

**1.标准参考物质**

**1.1 CRM的性质**

CRM可以是液体、气体或固体。不论其为何种物理状态,作为一个CRM,它应具备良好的稳定性、均匀性以及量值的准确性。

**1.2 CRM的定值**

CRM的定值常采用以下三类方式进行。

第一类是多个实验室协作定值,至少10个实验室重复分析所考虑的每个元素,对获得的数据进行单向的方差分析。在执行这类计划时,参与工作的实验室也可采用他们自己选定的方法,在证书上通常报告如下的统计参数:

Ac——元素含量的证书值。Ac一般是定值计划结果的总体平均值（或者是各实验室均值的平均值）;

V〔Ac〕——Ac的方差,证书值估价的精密度量度。间或精密度也根据置信区间来表达（置信度一般为95%）;

S c——各实验室采用其自己选择的方法时,实验室之间的标准偏差;

Src——各实验室采用其自己选择的方法时,实验室内部的标准偏差。

这种定值计划的另一种做法是参与工作的实验室均采用已经确定的广泛接受的同一方法,通常报告的统计参数是Ac、V〔Ac〕（定义同上）以及$\sigma_{LC}$、$\sigma_{rc}$。其中,

$\sigma_{LC}$——各实验室采用同一方法时,实验室之间的标准偏差。

$\sigma_{rc}$——各实验室采用同一方法时,实验室内部的标准偏差。

多个实验室的协作定值是研制CRM的最常用方式。

在一些特殊场合,当采用第一类方式不可能达到实验室间相一致的定值时,可采用第二类定值方式。某些权威实验室在对候选CRM进行分析,并接受一致的测定值,或者取考虑了分析方法的因素后的测定值作为证书值Ac时,就不进行分析结果的统计评价,元素证书值的不确定度系由化学因素确定。

46

第三类方式是在一个实验室采用权威方法进行鉴定，该方法是基于绝对测量原理建立的，具有高精度且其不确定度可用高置信度来表达。所报告的统计参数是权威实验室采用权威方法所测得的 Ac 和 σrc。

### 1.3 分析方法的验证

应用 CRM 测定分析方法的精密度和准确度有以下三个基本前提。第一、CRM 中元素的证书值是真值的最好估价。第二、当用多个实验室协作定值的 CRM 进行试验时，如果测定结果足够精密与准确，则可以认为它们与 CRM 鉴定证书上报告的参数的统计处理结果相符合。从统计意义上讲，它们属于 CRM 的特征分布状态。第三、实验室内部与实验室之间的标准偏差的关系，既适用于根据实验室协作定值的 CRM，也适用于根据权威实验室用权威方法定值的 CRM。

### 1.4 CRM 的使用

鉴于任何分析方法其误差是不可避免的，因此，有必要对每个 CRM 定义一个可接受的允许差，它在一定程度上取决于定值方式。应用 CRM 评价一个分析方法的准则为：

精密度：如果该分析方法用于分析某一 CRM 所得结果的精密度，与证书给出的实验室的精密度相一致，则可以认定该方法是足够精密的。

准确度：如果该分析方法用于分析某一 CRM，所得到的结果与证书值的偏差，不超出室内和室间的允许统计波动，则可认为该方法具有足够的准确度。

## 2. 用 CRM 验证分析方法的一般程序

### 2.1 CRM 的验证

CRM 在使用之前，必须对以下几项进行复核：

证书包括：证书值、精密度、鉴定方法、建议的最少称样量以及证书值的有效期（特别是对于较不稳定的 CRM）。

外观检查指：聚集状态，色泽情况及数量。

### 2.2 CRM 的分析

用所验证的方法分析 CRM 多次，分析次数（n）至少为 10 次。对分析方法的准确度作定期检查时，每次作双份测定就足够了，不过重复测定的总数应多于 10 次。

### 2.3 界外值

用一种界外值检验法，如 Dixon 检验法，来判断哪些结果属界外值，如果找出了界外值，研究产生界外值的物理原因。倘若过多结果是界外值，又不能认作是物理原因所引起的，那么要对分析步骤作再次研究。

### 2.4 均值和标准偏差

对舍弃界外值后保留的结果作如下的统计计算：

$$\overline{X} = \sum_{i=1}^{n} X_i / n = 均值$$

$$S_w = \sqrt{\sum_{i=1}^{n} \frac{(X_i - \overline{X})^2}{(n-1)}}$$

= 估计的室内标准偏差（即方法的精密度）。

### 2.5 精密度检验

如果室内标准偏差 Src 或 σrc 在证书中已给出，这步才能进行，否则要从化学角度来判断其精密度。

计算

$$F = \frac{S^2_w}{S^2_{rc}}$$

将 F 与 $F_0 = F_{0.95; n-1}$；DF₀（可从统计书上查得）进行比较，如果 CRM 的定值试验的自由度 DF₀ 在证书上未给出，取 $DF_0 = 60$。若 $F \leq F_0$，该方法是足够精密的；若 $F > F_0$，该方法不如该 CRM 的定值分析方法精密。

### 2.6 准确度检验

如果在证书上已给出室间标准偏差 SLc 或 σLc，则可作如下检验，设

$$|\overline{X} - Ac| \leq 2S_{Lc} \quad 或 \quad |\overline{X} - Ac| \leq 2\sigma_{Lc}$$

则说明该方法是足够准确的，否则，就认为它与参与定值计划实验室所采用的分析方法的准确度不一。

对于以下CRM，即根据多个实验室协作定值而未报告实验室之间标准偏差者；或者根据多个试验室协作采用一种广泛接受的与所验证的分析方法不同的方法定值者，或者根据一个权威实验室采用权威方法进行定值者，建议作如下的检验，如果

$$|X - A_c| \leq 4S_w$$

那么该方法是足够准确的。

### 3. 统计证明

对于所有统计检验 均采 用 $\alpha = 0.05$ 的显著性水平。

#### 3.1 重复测定次数（n）

重复测定次数直接影响室内精 密 度 $S_w$ 的估价。表1列出了在 $\alpha = 0.05$ 及各种 $\beta$ 值场合下，测定次数n与分析方法的标准 偏差 与 CRM的实验室内部标准偏差的 最大 比 值、$S/S_{rc}$ 之间的关系。如果试验方法 的标准偏差大于该比值乘以 $S_{rc}$ 的证书值，那么 $\beta$ 就是该方法将被接受的概率。例如 $\beta = 0.05$, $n = 10$，如果分析方法已经通过了F检验，那么标准偏差超过 $2.26S_{rc}$ 的几率 将 不 会 超过5%。

#### 3.2 报告了 $S_{LC}$ 的CRM

这有两个因素要考虑：证书值的不确定度及分析方法的不确定度。下式可以用作是否接受的准则。

$$|\overline{X} - A_c| \leq 2\sqrt{\sigma_c^2 + \sigma_D^2}$$

式中 $\sigma_c^2$ 是与证书值有关 的 不确定 度；$\sigma_D^2$ 是与所验证方法有关的不确定度，因 数2是 Student t—分布95%的概率的分位 数的 近似值。尽管有些粗糙，但是这种假定已被广泛接受，且已用于各种各样的ISO标准方法中。$\sigma_c^2$ 值取决于CRM的定值方法。

在采用几种方法执行多个实验室协作定值时：

表1. 当比值 $S_w/S_{rc}$ 大于表中所列数值时，重复测定次数（n）与分析方法将被接受的概率（β）的关系

| n−1 | α=0.05 | | | |
| --- | --- | --- | --- | --- |
| | β=0.01 | β=0.05 | β=0.1 | β=0.5 |
| 3 | 6.25 | 4.71 | 3.66 | 1.82 |
| 4 | 5.65 | 3.65 | 2.99 | 1.68 |
| 5 | 4.47 | 3.11 | 2.62 | 1.59 |
| 6 | 3.80 | 2.77 | 2.39 | 1.53 |
| 7 | 3.37 | 2.55 | 2.23 | 1.49 |
| 8 | 3.07 | 2.38 | 2.11 | 1.45 |
| 9 | 2.85 | 2.26 | 2.01 | 1.42 |
| 10 | 2.67 | 2.15 | 1.94 | 1.40 |
| 12 | 2.43 | 2.01 | 1.83 | 1.36 |
| 15 | 2.19 | 1.85 | 1.71 | 1.32 |
| 20 | 1.95 | 1.70 | 1.59 | 1.27 |
| 24 | 1.83 | 1.62 | 1.52 | 1.25 |
| 30 | 1.71 | 1.54 | 1.46 | 1.22 |
| 40 | 1.59 | 1.45 | 1.38 | 1.19 |
| 60 | 1.45 | 1.35 | 1.30 | 1.15 |
| 120 | 1.30 | 1.24 | 1.21 | 1.11 |

$$\sigma_c^2 = (S_{LC}^2 + S_{rc}^2/n_c)/N_c \quad (2)$$

在采用同一方法执行多个实验室协作定值时：

$$\sigma_c^2 = (\sigma_{LC}^2 + \sigma_{rc}^2/n_c)/N_c \quad (3)$$

这里，$N_c$ 是定值中提供分析数据的 实 验室数。$n_c$ 是 每个参与定值计划的实验室所进行重复测定的平均次数。在大多数多个实验室协作定值计划中，通常 $N_c > 10$，因此，与 $\sigma^2$ 相比，$\sigma_c^2$ 可以忽略不计。

$$\sigma_D^2 = \sigma_{LC}^2 + S_w^2/n \text{ 或 } S_{LC}^2 + S_w^2/n$$

$\sigma_D^2$ 取决于所用CRM的类型。对于 $n = 10$，则 $S_w^2/n$ 将变得更小。式（1）可简化为：

$$|\overline{X} - A_c| \leq 2S_{LC} \text{ 或 } |\overline{X} - A_c| \leq 2\sigma_{LC} \quad (4)$$

#### 3.3 未提供 $S_{LC}$ 的CRM

采用同一方法，以多个实验室协作定值的经验表明，只要 $\sigma_{LC}/\sigma_{rc} \leq 2$，此方法就可接受为标准方法。据此，可用 $2\sigma_{rc}$ 代 替 $S_c$，其可被接受的判据为：

$$|\overline{X} - A_c| \leq 4\sigma_{rc} \quad (5)$$

CCRMP应用多种方法以多个实验室 协作定值对135种元素和组分进行定值，且给出

48

了相应的允许差（$2S_{LC}$）和$S_{LC}/S_{rc}$比值，其中77种$S_{LC}/S_{rc}\leqslant2.0$，36种$S_{LC}/S_{rc}$或$\sigma_{LC}/\sigma_{rc}\leqslant2.5$。这些数据表明，对于数多多个实验室协作定值的方式来说，$S_{LC}/S_{rc}$或$\sigma_{LC}/\sigma_{rc}\leqslant2.0$的判别准则太粗糙，然而它却是合理的，特别是鉴于在CRM能够用作验证分析方法之前，实验室之间的标准偏差的某些估计是有必要的这一事实。据此，可用$2S_{rc}$代替$S_{LC}$，则可被接受的判据变成：

$$|\overline{X}-A_c|\leqslant4S_{rc} \qquad (6)$$

只要给出CRM的$\sigma_{rc}$或$S_{rc}$值，就可以用式（5）或式（6）检验分析方法的准确度。然而，运用式（7）同样可以很好地检验：

$$|\overline{X}-A_c|\leqslant4S_w \qquad (7)$$

如果方法的精密度经过检验，认为可以接受，即

$$F_0>F=\frac{S_w^2}{S_{rc}^2}\ 或\ \frac{S_w^2}{\sigma_{rc}^2}$$

在这种情况下，统计上$4S_w$等于$4\sigma_{rc}$法$4S_{rc}$。可见与式（5）、式（6）比较，使用（7）式的优点是：容易判别出较高精密度的方法具有较高的精密度。並且，在使用未给出$S_{rc}$或$\sigma_{rc}$的CRM时，可用它来检验准确度。

## 4.说明

对于某些没有给出证书值，仅提供了"最佳的"或"可使用的"数值，也未给出不确定度的CRM，这些值往往是根据很少的几次结果或基于相对偏差的一致的结果得出的结论。如标准岩石就是一个典型例子。如果使用这样的CRM及本文所建议的实验设计验证一个方法的准确度，那么，很好确定的只是元素或组分的量值V。其准确度的判据当然是$|\overline{X}-V|\leqslant4S_w$。

本文曾得到吴诚高级工程师的指点。在此谨致衷心的感谢。

**主要参考文献**

〔1〕 D. Oster, J. P. Caliste and A. Marschal, ISO／REMCO 72,（1981）

〔2〕 Y. Mashiko, K. Iizuki and S. Seaki, ISO／REMCO 86,（1982）

〔3〕 R. Sutarno and G. M. Faye, 《Talanta》,22,675（1975）

〔4〕 R. Sutarno and H. F. Steger 《Talanta》,32,439（1985）

鄂国强编译 （上海材料研究所四室）

的准确度。特别是辐射防护和辐射肿瘤治疗的测量准确度，为使我国的电离辐射剂量量值向国际基准量值靠扰，提供了有效的国际

保证。我实验室随时了解并掌握有关国际先进技术的信息和发展动态，提高我们的学术水平，而且可以提高我实验室在辐射剂量测量

# THE USE OF CERTIFIED REFERENCE MATERIALS IN THE VERIFICATION OF ANALYTICAL DATA AND METHODS*

R. SUTARNO and H. F. STEGER

Mineral Sciences Laboratories, CANMET, Energy, Mines and Resources Canada, Ottawa, Ontario, Canada

**Summary**—An experimental design is proposed for the verification of the accuracy and precision of an analytical method by its application to certified reference materials.

The importance of certified reference materials, CRMs, in analytical chemistry is widely accepted. There are at least 130 producers of CRMs of all types in the Western World alone.[1] If only compositional CRMs such as ores, rocks and metal alloys are considered, there are at least 39 producers, who together offer approximately 250 reference ores, concentrates, rocks, soils and cements and over 2400 metal alloys.

CRMs are widely applied for quality control of analytical data by their use for the direct calibration of methods and instruments, verification of the precision, accuracy and sensitivity of methods used in control analyses, and the development and evaluation of new or reference methods. In spite of this, no general guidelines for the use of CRMs have been published. To remedy this, REMCO, the Council Committee on Reference Materials of the International Organization for Standardization, initiated the preparation of such a guide in September 1983.[2]

In the present paper we are not concerned with the appropriateness or the mechanics of the use of CRMs for these applications, but only the statistical transfer of the stated uncertainty in the certified value(s) and the application of this uncertainty by the analyst.

The use of CRMs in the direct calibration of methods and instruments has been studied by Oster *et al.* (France)[3] and by Mashiko *et al.* (Japan)[4] and is currently being expanded by Eberhardt (U.S.A.)[5] as contributions to the forthcoming REMCO guide. The use of CRMs in the determination of precision and accuracy and the sensitivity of methods, established or new, may have greater significance for the readers of analytical journals dedicated to the improvement of methodology. Most issues of such journals include one or more papers where CRMs are applied in this manner. A comparison of such papers

readily shows that there is a variety of procedures for validating a modified or new method by applying it to CRMs. These vary from those based on statistical concepts to those based on subjective considerations. The latter approach has the potential danger that after the assessment with a CRM a user-analyst striving for high accuracy and precision may have doubts about his method that prove unfounded if the uncertainty in the certified values is taken into account.

To make analysts more aware of the concepts behind the use of CRMs, the Canadian Certified Reference Materials Project (CCRMP), on the basis of the experience of 13 years in the preparation of reference ores and related materials,[6,7] proposes the following experimental design for the use of CRMs in determination of the accuracy and precision of analytical methods†. The use of the proposed design should of course, do away with subjectivity.

## Nature of CRMs

CRMs can be liquid, gaseous, or solid (as pure or mixed powders or individual artifacts). Regardless of their physical state, CRMs are packaged either in individually certified units or, more commonly, in units which are all assumed to be representative of a larger lot. In the latter case, the between-unit homogeneity is generally acceptable for all but the most exacting analytical purposes. Some CRMs however cannot be made truly homogeneous; the use of such materials will not be considered in this paper.

## Certification

There are basically three types of analytical programmes for certification of a reference material. The first is by interlaboratory consensus,[8] a minimum of 10 laboratories analysing in replicate one or more units of the reference material for the elements of interest. A one-way analysis of variance is performed on the data. In one version of this programme the participating laboratories use the method of their

---

†This experimental design has also been submitted by CCRMP to REMCO.

choice. The following statistical parameters are then usually reported on the certification document.

$A_c$ = the certified value for the content of an element. $A_c$ is usually the overall mean of the results of the certification programme (or the mean of the laboratory means) but is sometimes some other central value such as the geometric mean or the median.

$V[A_c]$ = the variance of $A_c$, a measure of precision of the estimate of the consensus value. Alternatively, the precision may be expressed in terms of a confidence interval (usually 95%, which means that if the interlaboratory programme is repeated 100 times, it is expected that 95 of the resultant consensus means will fall within this confidence interval).

$S_{Lc}$ = the between-laboratories standard deviation when each laboratory uses the method of its choice.

$S_{rc}$ = the within-laboratory standard deviation when each laboratory uses the method of its choice.

In the second version of the programme, the participating laboratories use the same well-established widely-accepted analytical method. The statistical parameters usually reported are $A_c$, $V[A_c]$, defined as above, and $\sigma_{Lc}$ and $\sigma_{rc}$.

$\sigma_{Lc}$ = the between-laboratories standard deviation when all laboratories use the same method.

$\sigma_{rc}$ = the within-laboratory standard deviation when all laboratories use the same method.

Certification by interlaboratory consensus is the most frequently encountered type of programme.

The second type of programme is met in certain specialized fields where it is impossible or impractical to arrive at a consensus value by a full interlaboratory programme. A small number of "expert" laboratories analyse the candidate reference material and the certified value $A_c$ is arrived at by accepting the consensus value or taking into account "methodological" considerations.[9] There is no formal statistical evaluation of the analytical results. The uncertainty of the certified value of an element is based on chemical considerations.

In the third type of programme the certification is done by a single laboratory using a definitive method, i.e., a method that is based on first principles, has high precision and for which the limits of uncertainty can be stated with a high degree of confidence.[10] In addition to accepting that the definitive method is the best available, the user assumes that the method is applied without bias. This assumption is usually made when the producer laboratory is a national or industrial reference laboratory where "superior" resources are available, or when the analysis involves a certain "art", and specialized skill is required. The

statistical parameters reported are $A_c$ and $\sigma_{rc}$ of the definitive method as performed by the "expert" certifying laboratory. Some producers of reference materials also report an evaluation of any systematic error not included in $\sigma_{rc}$.

### Validation of analytical methods

The concept presented here, of using CRMs to determine the precision and accuracy of an analytical method, rests on three basic premises. First, that the certified value of an element in a reference material is the best estimate of the true value. Second, when tested with a reference material certified by interlaboratory consensus, an analytical method is validated if the results are sufficiently precise and accurate to qualify for inclusion in the calculation of the statistical parameters for the interlaboratory programme for the CRM. That is, in statistical terms, they belong to the distribution used to characterize the CRM. Third, the relationship between the within-laboratory and between-laboratories standard deviation found to hold for reference materials certified by interlaboratory consensus, would also apply to a reference material certified by an "expert" certifying laboratory using a definitive method, if that material were certified by interlaboratory consensus. This last assumption is necessary because very few users can expect to achieve the same accuracy and precision as an "expert" certifying laboratory. In principle a between-laboratories variation must exist for every method.

### Use of CRMs

All analytical methods have inherent errors. Random error is always present and may be accompanied by systematic error (bias). It is therefore highly unlikely that the result obtained by a method being tested will be exactly the same as the certified value of a CRM; an acceptable tolerance limit must therefore be defined for each CRM. This limit will depend partly on the mode of certification. The following criteria are proposed in assessing an analytical method by applying it to a CRM.

*Precision.* An analytical method is sufficiently precise if it produces results for a CRM that are statistically as precise as the within-laboratory precision stated in the certification document.

*Accuracy.* An analytical method is sufficiently accurate if it produces results for a CRM that do not differ from the certified value by more than can be accounted for by within- and between-laboratory statistical fluctuations.

### GENERAL PROCEDURES FOR APPLYING AN ANALYTICAL METHOD TO A CRM

### Validation of the CRM

Before a CRM is used, the following should be checked.

*Certificate*—certified value, precision, method of characterization, suggested sub-sample weight, and expiration date of certification (particularly for a relatively unstable CRM).

*Visual check*—segregation, discoloration and quantity.

## Analysis of the CRM

Perform replicate determinations (on separate sub-samples) by the analytical method to be tested; $n$ should be at least 10. For a periodic check of accuracy of an analytical method, duplicate analysis on each occasion is sufficient, but the total number of replicates should be at least 10.

## Outliers

Use an outlier test such as the Dixon test to decide whether any of the results may be regarded as outliers. If an outlier is found, investigate possible physical causes for it. If an excessive number of results are found to be outliers and no physical cause can be identified, reinvestigate the analytical procedure.

## Mean and standard deviation

Compute the following statistics for the results remaining after rejection of outliers.

$$\bar{X} = \sum_{i=1}^{n} X_i/n = \text{mean}$$

$$S_W = \sqrt{\sum_{i=1}^{n} \frac{(X_i - \bar{X})^2}{(n-1)}}$$

= estimated within-laboratory standard deviation, *i.e.*, precision of the method

## Check of precision

This can only be performed if $S_{rc}$ or $\sigma_{rc}$, the certified within-laboratory standard deviation, is given in the certificate, otherwise the precision should be judged by chemical considerations.

Compute

$$F = \frac{(S_W)^2}{(S_{rc})^2}$$

and compare $F$ with the value of $F_c = F_{0.95;n-1;DF_c}$ (obtainable from any statistics book). If the number of degrees of freedom $DF_c$ for the CRM certification tests is not given in the certificate, use $DF_c = 60$. If $F \leq F_c$, the analytical method is sufficiently precise; if $F > F_c$, the analytical method is not as precise as those used for certification of the reference material.

## Check of accuracy

If the between-laboratories standard deviation $S_{Lc}$ or $\sigma_{Lc}$ is reported for the CRM, *i.e.*, it was certified by interlaboratory consensus, then the following test should be used. If

$$|\bar{X} - A_c| \leq 2S_{Lc} \quad \text{or} \quad |\bar{X} - A_c| \leq 2\sigma_{Lc}$$

then the analytical method is sufficiently accurate. Otherwise it is not considered to be as accurate as those used by the laboratories participating in the certification programme.

For reference materials certified by interlaboratory consensus for which the between-laboratories standard deviation is not reported, or by a partial interlaboratory programme, or by interlaboratory consensus on the basis of one widely-accepted method that is not the same as the method being tested, or by

Table 1. Relation between the number of replicates ($n$) and the chance ($\beta$) that an analytical method will be accepted as valid, when the ratio of its standard deviation to the within-laboratories standard deviation of the CRM (at $\alpha = 0.05$) is greater than the tabulated value (calculated from data in *Handbook of Tables of Probability and Statistics*, 2nd Ed., p. 299, CRC, Cleveland, Ohio, 1968)

| | $\alpha = 0.05$ | | | |
|---|---|---|---|---|
| $n-1$ | $\beta = 0.01$ | $\beta = 0.05$ | $\beta = 0.1$ | $\beta = 0.5$ |
| 1 | 159.5 | 31.3 | 15.6 | 2.73 |
| 2 | 17.3 | 7.64 | 5.33 | 2.08 |
| 3 | 6.25 | 4.71 | 3.66 | 1.82 |
| 4 | 5.65 | 3.65 | 2.99 | 1.68 |
| 5 | 4.47 | 3.11 | 2.62 | 1.59 |
| 6 | 3.80 | 2.77 | 2.39 | 1.53 |
| 7 | 3.37 | 2.55 | 2.23 | 1.49 |
| 8 | 3.07 | 2.38 | 2.11 | 1.45 |
| 9 | 2.85 | 2.26 | 2.01 | 1.42 |
| 10 | 2.67 | 2.15 | 1.94 | 1.40 |
| 12 | 2.43 | 2.01 | 1.83 | 1.36 |
| 15 | 2.19 | 1.85 | 1.71 | 1.32 |
| 20 | 1.95 | 1.70 | 1.59 | 1.27 |
| 24 | 1.83 | 1.62 | 1.52 | 1.25 |
| 30 | 1.71 | 1.54 | 1.46 | 1.22 |
| 40 | 1.59 | 1.45 | 1.38 | 1.19 |
| 60 | 1.45 | 1.35 | 1.30 | 1.15 |
| 120 | 1.30 | 1.24 | 1.21 | 1.11 |

R. Sutarno and H. F. Steger

Table 2. Permissible tolerance ($2S_{Lc}$) values and $S_{Lc}/S_{rc}$ ratios for CCRMP reference materials

| RM | Certified species | $A_c$ | $S_{Lc}/S_{rc}$ | $2S_{Lc}$ |
|---|---|---|---|---|
| CT-1 | Tungsten, % | 1.04 | 1.3 | 0.066 |
| BH-1 | Tungsten, % | 0.422 | 1.6 | 0.0285 |
| TLG-1 | Tungstèn, % | 0.083 | 2.0 | 0.0140 |
| BL-1 | Uranium, % | 0.022 | 1.2 | 0.0016 |
| BL-2 | Uranium, % | 0.453 | 1.1 | 0.0137 |
| BL-3 | Uranium, % | 1.02 | 0.9 | 0.024 |
| BL-4 | Uranium, % | 0.173 | 1.2 | 0.0099 |
| BL-5 | Uranium, % | 7.09 | 1.0 | 0.152 |
| CCU-1 | Silver, $\mu g/g$ | 139 | 3.6 | 15.3 |
| CCU-1 | Alumina, % | 0.247 | 1.1 | 0.0217 |
| CCU-1 | Gold, $\mu g/g$ | 7.5 | 1.8 | 1.05 |
| CCU-1 | Copper, % | 24.71 | 1.3 | 0.231 |
| CCU-1 | Mercury, $\mu g/g$ | 61 | 2.0 | 7.8 |
| CCU-1 | Lead, % | 0.106 | 2.1 | 0.0236 |
| CCU-1 | Silica, % | 2.61 | 2.5 | 0.260 |
| CCU-1 | Zinc, % | 3.22 | 2.6 | 0.173 |
| CD-1 | Antimony, % | 3.57 | 2.3 | 0.152 |
| CD-1 | Arsenic, %. | 0.66 | 2.3 | 0.66 |
| CPB-1 | Silver, $\mu g/g$ | 62.6 | 1.9 | 26.4 |
| CPB-1 | Alumina, % | 0.28 | 1.8 | 0.039 |
| CPB-1 | Arsenic, % | 0.056 | 1.5 | 0.0136 |
| CPB-1 | Bismuth, % | 0.023 | 2.0 | 0.0040 |
| CPB-1 | Cadmium, % | 0.0143 | 1.5 | 0.0016 |
| CPB-1 | Copper, % | 0.254 | 1.8 | 0.0167 |
| CPB-1 | Iron, % | 8.43 | 2.5 | 0.251 |
| CPB-1 | Mercury, $\mu g/g$ | 5.15 | 1.4 | 1.59 |
| CPB-1 | Lead, % | 64.74 | 2.2 | 0.586 |
| CPB-1 | Sulphur, % | 17.8 | 2.2 | 0.41 |
| CPB-1 | Antimony, % | 0.36 | 2.4 | 0.045 |
| CPB-1 | Silica, % | 0.74 | 1.7 | 0.109 |
| CPB-1 | Zinc, % | 4.42 | 1.9 | 0.191 |
| CZN-1 | Silver, $\mu g/g$ | 93 | 2.4 | 10.2 |
| CZN-1 | Alumina, % | 0.25 | 2.2 | 0.043 |
| CZN-1 | Arsenic, % | 0.026 | 2.5 | 0.0089 |
| CZN-1 | Cadmium, % | 0.132 | 1.7 | 0.0085 |
| CZN-1 | Copper, % | 0.144 | 2.2 | 0.0138 |
| CZN-1 | Iron, % | 10.93 | 2.7 | 0.284 |
| CZN-1 | Mercury, $\mu g/g$ | 43 | 2.3 | 11.9 |
| CZN-1 | Manganese, % | 0.219 | 3.0 | 0.0258 |
| CZN-1 | Lead, % | 7.43 | 1.6 | 0.246 |
| CZN-1 | Sulphur, % | 30.2 | 2.8 | 0.78 |
| CZN-1 | Antimony, % | 0.052 | 1.9 | 0.0081 |
| CZN-1 | Zinc, % | 44.74 | 2.2 | 0.573 |
| DH-1A | Thorium, % | 0.091 | 2.7 | 0.0101 |
| DL-1A | Uranium, % | 0.0116 | 1.2 | 0.0013 |
| DL-1A | Thorium, % | 0.0076 | 2.0 | 0.0016 |
| HV-1 | Copper, % | 0.522 | 1.1 | 0.0201 |
| HV-1 | Molybdenum, % | 0.058 | 1.0 | 0.0047 |
| KC-1 | Lead, % | 6.87 | 2.0 | 0.222 |
| KC-1 | Copper, % | 0.112 | 2.1 | 0.0097 |
| KC-1 | Tin, % | 0.67 | 3.2 | 0.700 |
| KC-1 | Silver, % | 0.112 | 2.2 | 0.0064 |
| KC-1 | Zinc, % | 20.07 | 1.3 | 0.209 |
| MA-1 | Gold, $oz/ton$ | 0.519 | 0.7 | 0.0224 |
| MA-2 | Gold, $\mu g/g$ | 1.86 | 1.6 | 0.244 |
| MP-1 | Tin, % | 2.43 | 1.6 | 0.313 |
| MP-1 | Copper, % | 2.09 | 2.3 | 0.113 |
| MP-1 | Lead, % | 1.88 | 1.7 | 0.105 |
| MP-1 | Molybdenum, % | 0.014 | 0.9 | 0.0028 |
| MP-1 | Indium, % | 0.069 | 1.7 | 0.0067 |
| MP-1 | Bismuth, % | 0.024 | 1.9 | 0.0069 |
| MP-1 | Arsenic, % | 0.77 | 2.6 | 0.080 |
| MP-1 | Silver, $\mu g/g$ | 57.9 | 3.1 | 6.72 |
| MP-1 | Zinc, % | 15.90 | 1.4 | 0.236 |
| MP-1A | Zinc, % | 19.02 | 1.7 | 0.417 |
| MP-1A | Tin, % | 1.28 | 2.6 | 0.137 |
| MP-1A | Copper, % | 1.44 | 1.5 | 0.037 |
| MP-1A | Lead, % | 4.33 | 2.8 | 0.140 |

| RM | Certified species | $A_c$ | $S_{Lc}/S_{rc}$ | $2S_{Lc}$ |
|---|---|---|---|---|
| MP-1A | Molybdenum, % | 0.029 | 3.3 | 0.0044 |
| MP-1A | Indium, % | 0.033 | 2.2 | 0.0042 |
| MP-1A | Bismuth, % | 0.032 | 3.3 | 0.0062 |
| MP-1A | Arsenic, % | 0.84 | 2.1 | 0.070 |
| MP-1A | Silver, $\mu g/g$ | 69.7 | 2.5 | 6.94 |
| MP-2 | Silver, $\mu g/g$ | 4.9 | 2.3 | 1.0 |
| MP-2 | Bismuth, % | 0.246 | 3.8 | 0.0210 |
| MP-2 | Molybdenum, % | 0.281 | 3.5 | 0.0386 |
| MP-2 | Tungsten, % | 0.65 | 3.3 | 0.070 |
| MW-1 | Iron, % | 66.08 | 1.4 | 0.210 |
| MW-1 | Iron(II), % | 1.36 | 1.5 | 0.107 |
| MW-1 | Alumina, % | 0.30 | 2.7 | 0.042 |
| MW-1 | Silica, % | 4.60 | 2.5 | 0.216 |
| MW-1 | Lime, % | 0.053 | 1.2 | 0.0095 |
| MW-1 | Magnesia, % | 0.034 | 2.2 | 0.0079 |
| MW-1 | Phosphorus, % | 0.011 | 2.4 | 0.0031 |
| MW-1 | Potassium, % | 0.011 | 2.3 | 0.0024 |
| OKA-1 | Niobium, % | 0.37 | 3.7 | 0.047 |
| PD-1 | Lead, % | 2.75 | 3.0 | 0.287 |
| PD-1 | Arsenic, % | 0.77 | 2.2 | 0.118 |
| PD-1 | Mercury, $\mu g/g$ | 389 | 2.0 | 66.6 |
| PR-1 | Molybdenum, % | 0.594 | 3.1 | 0.0450 |
| PR-1 | Bismuth, % | 0.111 | 1.4 | 0.0102 |
| PR-1 | Iron, % | 1.244 | 1.9 | 0.0522 |
| PR-1 | Sulphur, % | 0.793 | 2.6 | 0.0522 |
| PTA-1 | Platinum, $oz/ton$ | 0.089 | 0.2 | 0.0065 |
| PTC-1 | Palladium, $oz/ton$ | 0.37 | 0.8 | 0.073 |
| PTC-1 | Platinum, $oz/ton$ | 0.087 | 0.7 | 0.0194 |
| PTC-1 | Rhodium, $oz/ton$ | 0.018 | 1.2 | 0.0038 |
| PTC-1 * | Gold, $oz/ton$ | 0.019 | 0.8 | 0.0083 |
| PTM-1 | Gold, $oz/ton$ | 0.052 | 1.0 | 0.0131 |
| PTM-1 | Palladium, $oz/ton$ | 0.24 | 1.1 | 0.051 |
| PTM-1 | Platinum, $oz/ton$ | 0.17 | 0.9 | 0.023 |
| RU-1 | Zinc, % | 2.24 | 2.4 | 0.084 |
| RU-1 | Copper, % | 0.85 | 1.8 | 0.035 |
| RU-1 | Iron, % | 24.40 | 2.2 | 0.322 |
| RU-1 | Sulphur, % | 21.71 | 2.5 | 0.454 |
| SL-1 | Silica, % | 35.73 | 1.7 | 0.466 |
| SL-1 | Lime, % | 37.48 | 2.5 | 0.773 |
| SL-1 | Magnesia, % | 12.27 | 2.4 | 0.686 |
| SL-1 | Alumina, % | 9.63 | 1.1 | 0.266 |
| SL-1 | Iron, % | 0.72 | 1.5 | 0.106 |
| SL-1 | Sulphur, % | 1.26 | 2.1 | 0.094 |
| SU-1 | Copper, % | 0.87 | 1.7 | 0.035 |
| SU-1 | Nickel, % | 1.51 | 1.1 | 0.053 |
| SU-1A | Nickel, % | 1.233 | 1.8 | 0.0327 |
| SU-1A | Copper, % | 0.967 | 1.3 | 0.0224 |
| SU-1A | Cobalt, % | 0.041 | 1.3 | 0.0039 |
| SU-1A | Platinum, $\mu g/g$ | 0.41 | 1.9 | 0.148 |
| SU-1A | Palladium, $\mu g/g$ | 0.371 | 1.7 | 0.062 |
| SU-1A | Silver, $\mu g/g$ | 4.3 | 1.3 | 0.76 |
| TAN-1 | Tantalum | 0.236 | 1.1 | 0.0217 |
| UM-1 | Copper, % | 0.43 | 1.7 | 0.021 |
| UM-1 | Nickel, % | 0.88 | 1.5 | 0.038 |
| UM-1 | Cobalt, % | 0.035 | 1.9 | 0.0039 |
| SCH-1 | Iron, % | 60.73 | 2.2 | 0.396 |
| SCH-1 | Silicon, % | 3.78 | 2.0 | 0.173 |
| SCH-1 | Aluminium, % | 0.509 | 1.6 | 0.0390 |
| SCH-1 | Calcium, % | 0.029 | 2.9 | 0.0097 |
| SCH-1 | Magnesium, % | 0.020 | 1.9 | 0.0053 |
| SCH-1 | Manganese, % | 0.777 | 1.9 | 0.0377 |
| SCH-1 | Titanium, % | 0.031 | 1.7 | 0.0072 |
| SCH-1 | Sulphur, % | 0.007 | 2.0 | 0.0026 |
| SCH-1 | Phosphorus, % | 0.054 | 1.2 | 0.0102 |
| SCH-1 | Sodium, % | 0.019 | 2.6 | 0.0090 |
| SCH-1 | Potassium, % | 0.026 | 2.6 | 0.0096 |

one "expert" laboratory using a definitive method, the following test is suggested, If

$$|\bar{X} - A_c| \leqslant 4S_w$$

then the analytical method being tested is sufficiently accurate.

## STATISTICAL JUSTIFICATION

A significance level of $\alpha = 0.05$ is used for all statistical tests.

### Number of replicate determinations, n

The number of replicate determinations to be performed directly affects the precision of the estimate of the within-laboratory precision, $S_w$. Table 1 shows the relation between $n$ and the maximum ratio of the standard deviation of the analytical method and the within-laboratories standard deviation for the CRM, $S_w/S_{rc}$, at various values of $\beta$, and $\alpha = 0.05$, $\beta$ being the probability that the method will be accepted if the standard deviation is greater than the ratio times the certified value of $S_{rc}$. For example, at $\beta = 0.05$ for $n = 10$, if the analytical method has passed the appropriate $F$-test, there is no more than a 5% chance that it has a standard deviation in excess of $2.26S_{rc}$.

### CRMs for which $S_{Lc}$ is reported

There are two factors to consider: the uncertainty of the certified value and the uncertainty of the analytical method. The following formula can be used as the criterion for acceptance.

$$|\bar{X} - A_c| \leqslant 2\sqrt{\sigma_C^2 + \sigma_D^2} \tag{1}$$

where $\sigma_C^2$ is the uncertainty associated with the certified value and $\sigma_D^2$ is the uncertainty associated with the analytical method being examined. The factor 2 is an approximation for the fractile of *Student's* $t$-distribution for 95% probability. This assumption, although somewhat crude, has been widely accepted and used in various ISO standard methods. The value of $\sigma_C^2$ depends on the method of certification of the reference material.

For certification by interlaboratory consensus from use of several methods:

$$\sigma_C^2 = (S_{Lc}^2 + S_{rc}^2/n_c)/N_c \tag{2}$$

For certification by interlaboratory consensus by use of one method:

$$\sigma_C^2 = (\sigma_{Lc}^2 + \sigma_{rc}^2/n_c)/N_c \tag{3}$$

where $N_c$ is the number of laboratories which produced results used in certification of the reference material, and $n_c$ is the average number of replicate determinations performed by each participating laboratory for the certification programme. For most certifications by interlaboratory consensus, $N_c$ is usually $> 10$, so $\sigma_C^2$ is small enough in relation to $\sigma_D^2$ to be neglected, and then

$$\sigma_D^2 = \sigma_{Lc}^2 + S_w^2/n \quad \text{or} \quad S_{Lc}^2 + S_w^2/n$$

depending on the type of CRM used. For $n = 10$, $S_w^2/n$ becomes small and equation (1) can be simplified to

$$|\bar{X} - A_c| \leqslant 2S_{Lc} \quad \text{or} \quad |X - \bar{A}_c| \leqslant 2\sigma_{Lc} \tag{4}$$

### CRMs for which $S_{Lc}$ is not available

Experience in interlaboratory consensus programmes based on use of one method, by ISO Technical Committee 102 on iron ores, shows that an analytical method is considered to be well standardized if $\sigma_{Lc}/\sigma_{rc} \leqslant 2$. On this basis, $S_{Lc}$ can be replaced by $2\sigma_{rc}$ and the criterion for acceptance is:

$$|\bar{X} - A_c| \leqslant 4\sigma_{rc} \tag{5}$$

Table 2 reports the value of $S_{Lc}/S_{rc}$ for 135 elements and constituents that have been certified by CCRMP by interlaboratory consensus based on use of multiple methods. Of these, 77 show $S_{Lc}/S_{rc} \leqslant 2.0$ and all but 22 show $S_{Lc}/S_{rc}$ or $\sigma_{Lc}/\sigma_{rc} \leqslant 2.5$. These data suggest that the criterion $S_{Lc}/S_{rc}$ or $\sigma_{Lc}/\sigma_{rc} \leqslant 2.0$ is too harsh for most interlaboratory consensus programmes, but nevertheless is reasonable, particularly in view of the fact that some estimate of the between-laboratories standard deviation is required before the CRM can be used for validation of an analytical method. On this premise, $S_{Lc}$ can be replaced by $2S_{rc}$ and the criterion for acceptance becomes:

$$|\bar{X} - A_c| \leqslant 4S_{rc} \tag{6}$$

An analytical method can be verified for accuracy by use of equation (5) or (6) provided the value of $\sigma_{rc}$ or $S_{rc}$ for the CRM is reported. We propose, however, that the accuracy of the method can equally well be judged by:

$$|\bar{X} - A_c| \leqslant 4S_w \tag{7}$$

if the precision of the method can be verified as stated above, i.e.,

$$F_0 > F = \frac{S_w^2}{S_{rc}^2} \quad \text{or} \quad \frac{S_w^2}{\sigma_{rc}^2}$$

in which case $4S_w$ is statistically equal to $4\sigma_{rc}$ or $4S_{rc}$, i.e., the upper limit of $|\bar{X} - A_c|$ for acceptance of the method is not significantly raised. The advantage in the use of equation (7) is that, to be accepted, a method with higher precision must also have greater accuracy, i.e., give a lower value of $|\bar{X} - A_c|$, than would be necessary if equation (5) or (6) were used as criterion.

In use of CRMs for which $\sigma_{rc}$ or $S_{rc}$ is not reported, we propose the use of equation (7) to test accuracy. The acceptability of the value of $S_w$ must be decided by the analyst on chemical grounds.

## EXAMPLES OF USE OF CRMs

Table 3 summarizes the results of the validation of some analytical methods by their application to some

Table 3. Validation of analytical methods

| CRM | DL-1a | SU-1a | SCH-1 | BL-2a |
|---|---|---|---|---|
| Element | Th | Co | Na | U |
| Certification | Interlaboratory, multiple methods | Interlaboratory, multiple methods | Interlaboratory, single method | Definitive method, volumetric umpire |
| $A_c$ | 0.0076% | 0.041% | 0.0186% | 0.426% |
| $S_{Lc}$, $\sigma_{Lc}$ | 0.0008 | 0.0020 | 0.0045 | — |
| $S_{rc}$, $\sigma_{rc}$ | 0.0004 | 0.0013 | 0.0017 | 0.0015 |
| Method | Colorimetry | Atomic-absorption | Same as above | Neutron activation |
| $\bar{X}$ | 0.0080% | 0.0474% . | 0.0201% | 0.423% |
| $n$ | 10 | 10 | 10 | 45 |
| $S_w$ | 0.0005 | 0.0005 | 0.0006 | 0.0033 |
| Precision check | $F = 1.563$ $F_{0.95,9,60} = 2.04$ —accepted | $F = 0.148$ $F_{0.95,9,60} = 2.04$ —accepted | $F = 0.125$ $F_{0.95,9,60} = 2.04$ —accepted | $F = 4.84$ $F_{0.95,44,24} = 1.88$ —failed |
| Accuracy check $\lvert \bar{X} - A_c \rvert$ $2S_{Lc}$ or $2\sigma_{Lc}$ or $4S_w$ | 0.0004 $2S_{Lc} = 0.0016$ —accepted | 0.0064 $2S_{Lc} = 0.0040$ —failed | 0.0015 $2\sigma_{Lc} = 0.0090$ —accepted | 0.003 $4S_w = 0.013$ —accepted |

CRMs from the CCRMP. The examples are taken from CCRMP certification programmes. For DL-1a,[12] SU-1a[13] and SCH-1,[14] the method illustrated was used by one laboratory participating in the inter-laboratory programme. In all three instances, the method is capable of acceptable precision in comparison with the overall results of the interlaboratory programme. The atomic-absorption method used to determine cobalt in SU-1a, as used by a particular laboratory, does not give acceptable accuracy. Indeed the results from this laboratory were designated as outliers and rejected for use in the interlaboratory programme.

BL-2a[15] was certified for uranium content solely on the basis of the results obtained by the volumetric umpire method performed at the Canada Centre for Mineral and Energy Technology. The neutron-activation method being validated was applied by a commercial laboratory to confirm the between-unit homogeneity of BL-2a. A precision check for this type of CRM is not recommended but has been performed for illustrative purposes. A definitive method is sufficiently precise for almost all other methods to fail the precision check. The acceptance of the accuracy is contingent on a value of $S_w = 0.0033$ being acceptable to the experimenter.

Caveat

For certain reference materials internationally available, no certified values are given, and only "best" or "usable" values are supplied, for which uncertainties are not reported. These values are often based on very few results or on results in relatively poor agreement. The reference rocks available from many sources are a typical example. The validation of the accuracy of an analytical method, by use of such reference materials and the experimental design proposed here, should be performed with caution, and only for elements and constituents for which the reported value, $V$, is well defined. The criterion of accuracy is then, of course, $\lvert \bar{X} - V \rvert \leqslant 4S_w$.

REFERENCES

1. ISO Directory of Certified Reference Materials, 1st Ed., 1982.
2. Ninth Meeting of REMCO, Geneva, 12–15 September 1983.
3. D. Oster, J. P. Caliste and A. Marschal, ISO/REMCO 72, 1981.
4. Y. Mashiko, K. Iizuki and S. Seaki, ISO/REMCO 86, 1982.
5. K. R. Eberhardt, personal communication.
6. H. F. Steger, CANMET Report 83-3E, Energy, Mines and Resources Canada, Ottawa, 1983.
7. R. Sutarno and G. M. Faye, Talanta, 1975, 22, 675.
8. H. Marchandise and E. Colinet, Z. Anal. Chem., 1983, 316, 669.
9. J. P. Cali and W. P. Reed, in Accuracy in Trace Analysis: Sampling, Sample Handling and Analysis, P. D. Lafleur (ed.), Vol. I, p. 56. U.S. Department of Commerce, Washington D.C., 1976.
10. Idem, op. cit., p. 47.
11. H. F. Steger and W. S. Bowman, CANMET Report 80-10E, Energy, Mines and Resources Canada, Ottawa, 1980.
12. Idem, CANMET Report 80-9E, Energy, Mines and Resources Canada, Ottawa, 1980.
13. R. Sutarno, D. J. Charette, W. S. Bowman and G. H. Faye, CANMET Report 78-5, Energy, Mines and Resources Canada, Ottawa, 1978.
14. H. F. Steger, W. S. Bowman, G. Zechanowitsch and R. Sutarno, CANMET Report 82-6E, Energy, Mines and Resources Canada, Ottawa, 1982.